Author:
Orford, Scott

Title:
Valuing the built environment : a GIS approach to the hedonic modelling of housing markets.

Valuing the Built Environment:

A GIS Approach to the Hedonic Modelling of Housing Markets

Scott Orford

A thesis submitted to the University of Bristol in accordance with the requirements of the degree of Ph.D. in the Faculty of Social Science, Department of Geography.

September 1997

Abstract

The valuation of the built environment has been a traditional concern of geographers. A particular interest has been the way in which the value of locational externalities are incorporated into house prices through housing market dynamics. However, much of the previous research into this process has been of North American origin, despite the fact that house prices, and property valuations in general, have become a major part of British life. This research aims to begin to rectify this shortfall by studying the spatial dynamics of the Cardiff Housing Market. Implicit in this research is an attempt to move towards a valuation of locational externalities at the micro-scale.

The research employs two distinct method of analysis. Firstly, ARC / INFO GIS is used to construct a context-sensitive GIS of the Cardiff housing market. An important aspect of this GIS is the use of Ordnance Survey's ADDRESS-POINT product to geo-reference individual properties to a resolution of 0.1 metre. Several large and complex socio-economic and property related datasets were then attached to this coverage, including house price survey data, local taxation data, and data from a Housing Condition Survey of one in five dwellings in the central area of Cardiff. This GIS is one of the most comprehensive constructed for any city, and is relatively unique in this kind of research.

The second method of analysis employs the hedonic pricing technique to impute monetary values for the implicit attributes of housing. An important part of the research is an investigation into the specification of the hedonic house price function. The traditional specification is essentially aspatial, and does not take into account the spatial nature the data, and thus the spatial dynamics of the housing market that generates it. To rectify this, three different specifications of the hedonic house price function are investigated: the traditional specification, the spatial parameter drift specification and the multi-level specification. The research concludes that the multi-level specification is best at modelling the spatial heterogeneity and spatial dependence inherent in housing market data. The results from this modelling show that the valuation of locational externalities are intimately bound up with the attributes of the housing stock and the characteristics of the resident households, resulting in a complex juxtaposition of positive and negative valuations of location at the local level.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree.

Any view expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

SIGNED: S. Orford DATE: 25/09/97

To my parents

Acknowledgements

There are several people whom I would like to thank for helping me with this research. Firstly, my supervisor Professor Paul Longley, who has helped me throughout this research, and particularly for the time and trouble he has taken to read and re-read my draft chapters. I would also like to thank Mr Edward Thomas for supporting me during my time as a Computer Officer in the department, especially during the last month for allowing me time to finish off. I would also like to acknowledge the help of Dr Gary Higgs in the construction of the GIS, and in particular in the manual matching of ADDRESS-POINT and the Council tax register during the (long) summer of 1994. I would also like to thank Mr Simon Godden for help with some of the art work. I am also indebted to the numerous Estate Agents and Building Societies for supplying me with the necessary information free of charge. Finally, I would like to thank the University of Bristol, the Geography Department, and the ESRC for helping to fund this research.

On a more personal note, I would like to thank all my friends and colleagues in the Geography Department. A special thanks goes to Carmel Thomason for her warmth and friendship over the past three years. It's been great fun. Another big thank you goes to Sandy Holland and her entertaining emails from America. Thanks also to Mike Bullock for all those holidays to Italy with Franca and Co. (and not forgetting Auntie Jaq.). I would also like to thank my flat mates Richard Harris, Jo Southworth, Liz Parsons and Jon Anderson for putting up with me. Finally, and most importantly, I would like to thank my family for always encouraging me and supporting me in whatever I do.

Concise Table of Contents

Table of Contents

Chapter One
Introduction

Chapter Two
The Hedonic House Price Function

Chapter Three
Housing Attributes and Spatial Data

Chapter Four
Cardiff Case Study and Data Sources

Chapter Five
Constructing a Context-Sensitive GIS

Chapter Six
GIS and Exploratory Data Analysis

Chapter Seven
Modelling the Spatial Dynamics of the Cardiff Housing Market

Chapter Eight
Towards a Valuation of Locational Externalities

Chapter Nine
Conclusions

List of Illustrations

Chapter Two
The Hedonic House Price Function

Chapter Four
Cardiff Case Study and Data Sources

Chapter Five
Constructing a Context-Sensitive GIS

Chapter Six
GIS and Exploratory Data Analysis

Chapter Seven
Modelling the Spatial Dynamics of the Cardiff Housing Market

Chapter Eight
Towards a Valuation of Locational Externalities

List of Tables

Chapter Seven
Modelling the Spatial Dynamics of the Cardiff Housing Market

Chapter Eight
Towards a Valuation of Locational Externalities

Chapter One
Introduction

Section 1.1 Introduction

The valuation of the built environment has long been a traditional concern of geographers. Over the years, numerous conceptual, theoretical and empirical studies have attempted to formulate, model and quantify how the built environment is valued by the people who live there. This research aims to complement and extend some of this work by investigating how the built environment of is valued through an urban housing market. More specifically, the principal aim of this research is to move towards a valuation of locational externalities, by modelling housing market dynamics. This can be achieved by using the method of hedonic pricing. This is an economic technique used for estimating the monetary value of attributes of complex commodities. These attributes do not have directly observable market prices, but when totalled together, their values equal the market price of the commodity. Within this context, the price of a house can be regarded as the sum of the implicit prices of its attributes. Since location is an integral attribute of a house, its value can be estimated using an hedonic house price function. This function relates house price to housing attributes, with the resulting parameter estimates corresponding to the implicit prices of these attributes.

Hedonic house price research has become well established over the past three decades. However, this research has tended to be of North American orientation, with very little comparative work done in the UK. Nevertheless, in recent years, the valuation of the built environment has become increasingly important within the UK. The 1980s saw owner occupation grow to two thirds of all tenure types, whilst house prices increased at an unprecedented rate, before slumping steeply in the early 1990s (Dorling, 1995). The 1990s also saw the introduction of the council tax as a method of local taxation, based upon domestic property values. Some of these issues will be discussed in greater detail in subsequent chapters, and they highlight the importance of house prices and the built environment, and the growing need to study their interaction within a UK context (Longley et al, 1994).

The aim of this research is to begin to rectify this shortfall in UK based research by using the hedonic house price function to value locational externalities within the city of Cardiff, Wales. As such, the research is divided into two main parts. The first part will investigate and evaluate different specifications of the hedonic house price function, by modelling the spatial dynamics of the Cardiff housing market. The conclusions from this will then inform the second part of the study, which aims to value the locational externalities that impact upon house prices within a specified central area of Cardiff. Before this can be achieved, however, it is necessary to have some understanding of how a housing market operates, and how this can affect property prices.

There are many theories of how housing markets operate, but since this study is primarily interested in the owner occupied sector of a western capitalist housing market, the literature reviewed will reflect this concern. Moreover, hedonic house price research has its origins within the location and landuse theories formulated during the 1960s. As such, hedonic house price research is intimately bound up with the micro-economic literature of housing markets and residential location. Therefore, the remainder of this chapter is devoted to the micro-economic theories of housing markets and residential location, and the extent to which hedonic house price research has departed from these formulations. This review is necessarily discursive since the economic underpinnings of hedonic house price theory is explained in detail in *Chapter Two.*

Therefore, this chapter is divided into six sections. Section two provides a synopsis of each chapter, which will also provide a better understanding of the research aims. Section three will briefly discuss the theoretical and conceptual consideration of housing, whilst section four describes the micro-economic theory of the housing market, and the trade-off model of residential location. Section five explores the theoretical underpinnings of hedonic house price research, and how this departs from the micro-economic theory, and finally, section six concludes the chapter.

Section 1.2 Thesis Structure

Chapter Two of this thesis will place the concept of hedonic analysis into context. In particular, the specification of the hedonic house price function will be investigated, with specific emphasis upon the incorporation of space into the function. This is to ameliorate problems encountered by previous hedonic house price studies when attempting to model spatial data.

Chapter Three of this thesis will discuss the theory underlying the role of housing attributes and develop a critique of previous studies. This will include an examination of the concept and measurement of locational externalities. The chapter will also examine the problems associated with generating, manipulating and storing spatial data within the context of a Geographic Information System.

Chapter Four will set out the empirical research aims in full, referring to the discussions in *Chapters Two* and *Three*. These include constructing a context-sensitive GIS to act as a medium for the research, and examining the socio-economic datasets that will enable the built environment to be valued. *Chapter Five* will then describe the construction of the GIS and the integration of the datasets, whilst *Chapter Six* will describe how locational externalities were generated using the GIS, and how preliminary hedonic models were built in an investigatory capacity.

Chapters Seven and *Eight* will present the results of the research. *Chapter Seven* will investigate the spatial dynamics of the Cardiff housing market. In particular, it will attempt to evaluate the success of each of the hedonic house price specifications in modelling the spatial structures of the housing market. *Chapter Eight* will then evaluate the impact of specific locational externalities upon property price, and generate valuation maps of the geography of externality effects. *Chapter Nine* will then conclude the research, and discuss its implications.

Section 1.3 Housing and the Housing Market

1.3.1 Introduction

The aim of this section is to briefly review the concept of housing as a commodity, and some of the factors that influence its supply and demand on the housing market. Implicit in this is the concept of housing possessing both a use value and an exchange value. The differences between these valuations are important, and their effect upon the supply and demand of housing are briefly explained. The section then concludes with a brief description of the factors that influence a household's decision to move.

1.3.2 Housing as a Commodity

'It is fixed in geographic space, it changes hands infrequently, it is a commodity which we cannot do without, and it is a form of stored wealth which is subject to speculative activities in the market ... In addition, [it] has various forms of value to the user and above all it is the point from which the user relates to every other aspect of the urban scene' (Harvey, 1972; pp 16)

Housing is unlike most other commodities. It is a complex package of goods and services that extends well beyond the shelter provided by the dwelling itself. Housing is also a primary determinant of personal security, autonomy, comfort, well being and status, and the ownership of housing itself structures access to other scarce resources, such as educational, medical, financial and leisure facilities (Knox, 1995). As such, housing has been viewed as a 'composite demand for a flow of services embodying a variable mix of characteristics' (Maclennan, 1982, pp. 41) - a multi-dimensional commodity. It is so intimately bound up with the lives of individuals that only one of its kind is usually consumed by a household at any time (Muth & Goodman, 1989).

It is typical to talk of a household purchasing packages or bundles of housing services (Bourne, 1981), which vary between housing types and housing markets. However, defining what a particular bundle actually is can be complicated. In addition, housing has a number of relatively unique attributes. It has a fixed location, a long durability, and a limited adaptability in response to changing demands. Housing stock is complex and diverse and is sensitive to changes that are external to the local market. Housing is also subjected to a multitude of institutional regulations imposed by government.

1.3.3 Supply and Demand of Housing

When discussing issues of supply and demand of housing, it is important to make the distinction between use value and exchange value. Use value generally refers to the net utility supplied by the bundle of housing services, whilst the exchange value of a property is the capital value it can realise in a competitive housing market. Although the use value of a property is a major determinant of its exchange value, this will also be influenced by the property's potential for increasing capital gain, since the purchase of housing stock is often the largest and only source of a household's accumulated savings (Muth & Goodman, 1989). Thus, two different housing markets can be identified. One deals with the supply and demand of bundles of housing services, whilst the other deals with an asset that can be termed housing stock. Although these two markets are conceptually different, they are integrated, since the majority of houses will offer similar services, such as a water supply.

Therefore, housing demand is a reflection of both its use value for consumption or occupancy purposes and its exchange value as an investment good. Housing demand tends to vary between income and racial groups and at different stages of the family life-cycle. Other influential factors include migration, immigration and changes in tax rates and taxation policies, particularly mortgage interest rates. Demand for housing has been of central interest to economic theorists and has been contextualised in micro-economic model of land use (Alonso, 1964) and residential location (Evans, 1973) - see subsequent sections. However, as will be explained, these demand led models have been severely criticised for ignoring the supply of housing.

The majority of properties supplied on the housing market come from the existing housing stock. Only a small proportion of the supply comes from newly constructed properties (Bourne, 1981). A substantial proportion of new supplies from existing stock arise through the subdivision of property and the conversion of non-residential buildings to dwelling uses. Even more occur through the death of a household, through the move of an existing household to shared accommodation or by a move outside the city. Supplies of housing stock may be ended by demolition or conversion to a non-residential use, or even merger by knocking together two or more dwellings. (Knox, 1995). Therefore, housing supply is a complex phenomenon with new supply and existing supply requiring separate, but interdependent analysis (Maclennan, 1982). Furthermore, the supply of housing will experience time lags between the decision to supply housing services and these housing services coming onto the market. The rate at which these new supplies enter the market in the short term is sensitive to house price changes and fluctuations in interest rates. This is particularly so for new constructions in which the price and the availability of land, planning controls and the provision of infrastructure have important influences on decisions regarding the location and timing of a development (Muth & Goodman, 1989).

1.3.4 Factors Influencing the Decisions to Move

The household is initially assumed to be receiving a given utility in their present dwelling. For movement to be considered, a minimum threshold of housing dissatisfaction must be perceived. This may occur over a period of time as the household may recognise its existing mismatch of housing attributes and household activities. The decision to enter the housing market and evaluate alternative housing opportunities may be triggered by a variety of factors. These may include increases in income, changes in family size, household formulation, or relative price changes across the market. Knox (1995) has made the distinction between voluntary and involuntary moves. Voluntary movements may be initiated by dissatisfaction with dwelling and garden space, housing repair costs and style obsolescence, as well as complaints about the neighbourhood. Reasons for forced moves include marriage, divorce, a death in the family, retirement, ill-health, and employment changes. However, almost two thirds of household movement is due to changes in the family life-cycle, and their perceived space requirements (Short, 1982).

In recent years, however, another factor has emerged that has had an important influence on the propensity to move in the UK; negative equity. This occurs when the market price of a house is less than the mortgage secured upon the property, and this became a widespread problem at the end of the 1980s and 1990s, when house prices slumped dramatically in many parts of the UK (Dorling, 1995). Negative equity means that the household is liable to cover the additional money secured on the property when it is sold, and having to find this money prevented many households from being able to move in the early 1990s. Negative equity particularly affected first time buyers, young buyers and less affluent buyers, since these were more likely to take out relatively larger loans and then have less ability to pay them back via earnings, inheritance and other assets. Low levels of equity can also deter households from moving.

Section 1.4 The Micro-Economic Theory of Housing Markets

1.4.1 Introduction

The 1960s and 1970s witnessed a proliferation of micro-economic theories and mathematical models of housing markets, residential location and landuse in both the UK and USA (e.g. Alonso, 1964; Muth, 1969; Batty, 1976). These formulated housing market dynamics in purely economic terms, based upon the theory of the firm and consumer behaviour. It will be explained that a key element to these formulations was the concept of a housing market in perfect equilibrium, functioning under Pareto Optimum conditions. These micro-economic theories subsequently underpinned neo-classical approaches to residential location, and in particular, the trade-off model of residential location. The trade-off model is perhaps one of the most influential economic models of residential location within hedonic house price theory. However, as will be discussed in the next section, hedonic house price theory has since abandoned much of the micro-economic theory concerning perfectly functioning housing markets and Pareto Optimum conditionality, in favour of a segmented housing market in disequilibrium. Therefore this section will briefly discuss the micro-economic theories of housing markets and residential location, before examining how this has been reformulated under hedonic house price theory.

1.4.2 The Perfectly Competitive Housing Market

The owner-occupied housing market is primarily an economic market set within a political framework for the purpose of exchanging housing services. In economic theory, the role of the market is to allocate scarce resources in an efficient manner so as to maximise output while minimizing cost, using price as the allocation mechanism. The most precise interpretations of this conceptualisation of the housing market derives primarily from the micro-economics literature. Maclennan (1982) identifies nine assumptions that define a set of conditions sufficient for the existence of a perfectly competitive housing market. These focus on the behaviour of individual producers and consumers, and views the matching of households to housing units as essentially an assignment problem (Bourne, 1981). The allocation proceeds as to achieve a market clearing solution; one in which all housing units are allocated and all households are accommodated, in the most efficient way. The assignment is also optimal in the sense that no household could be made better off with a different assignment without making another household worse off . This is known as Pareto Optimum conditions, and is a source of contention in the hedonic house price literature. As be discussed is a later section, hedonic house price research argues against Pareto Optimum equilibrium conditions in favour of a segmented housing market in disequilibrium. However, before this can be explored in more detail, it is necessary to set out the conditions under which a perfectly functioning housing market is said to operate.

Following Maclennan (1982, pp. 36), Pareto Optimum conditions can be achieved under the following nine assumptions:

1. There are many buyers and sellers
2. In relation to the aggregate volume of transactions the sales or purchases of each house are insignificant
3. There is no collusion amongst or between buyers and sellers
4. There is free entry into and exit from the market for both consumers and producers
5. Consumers have continuous, transitive and established preferences over a wide range of alternative choices of housing and non-housing goods

6. Consumers and producers possess both perfect knowledge with respect to prevailing prices and current bids and perfect foresight with respect to future prices and future bids.
7. Consumers maximise total utility[1] whilst producers maximise total profits.
8. There are no artificial restrictions placed on the demands for supplies and prices of housing services and the resources used to produce housing service. For instance, house purchases are not constrained by finance rationing or the non-availability of preferred housing choices.
9. The market is assumed to be in equilibrium

It can be seen that these assumptions are extremely idealised, and as such, easily critiqued. For instance, the abstracted assumptions of perfect competition and rational buyers and sellers are frequently cited (Ball, 1985). However, a major source of criticism of the micro-economic theories of housing markets has been the disregard of the supply of housing (ibid.). Compared to demand, very little micro-economic work has been done on the supply of housing, especially in the short-run (Muth and Goodman, 1989). For instance, in the micro-economic supply model developed by Muth (1969), a supplier has perfect information regarding present and future house price changes, and is assumed to be a price-taking profit maximiser. This allows the precise output level of housing to be identified deductively. This model has been used to formulate theoretical specifications for the estimation of price elasticity of supply of housing both in the short-run and long-run and for the elasticity of substitution between land and non-land inputs to housing supply. However, the durability of housing, and the difficulty of adapting existing stock to changes in demand has been ignored, even though these will effect long- and short-run housing market equilibriums.

1.4.3 The Neo-Classical Approach to Residential Location

1.4.3.1 Introduction

Concurrent with the formulation of the micro-economic theories of housing markets was the development of new approaches to residential location. Under the auspicious title of new urban economics, these neo-classical approaches were underpinned by similar micro-economic assumptions, and used comparative-static utility maximisation to deduced urban rent gradients

[1] See *Chapter Two*, section one for a discussion on housing utility.

and individual household demand functions for housing space and city centre access. By the mid-1970s these models had coalesced into a general theory of residential location known as the trade-off model. The remainder of this section will discuss the concept of the trade-off model and the implications that it has had for hedonic house price research.

1.4.3.2 The Trade-off Model

The trade-off model (Bassett & Short, 1980), or 'access-space' trade-off model (Maclennan, 1982), describes how households trade-off travel costs to the city centre, against housing costs in an attempt to maximise utility subject to an overall budget constraint. The theoretical background of the trade-off model was developed in two stages (Anas & Dendrinos, 1976). The basic models were developed during the 1960s, principally by Alonso (1964), Beckmann (1968) and Muth (1969), and were based upon the micro-economic theory of housing markets operating under Pareto Optimum conditions. These initial models were elaborated during the 1970s by economists such as Evans (1973), Mills (1972) and MacDonald (1979). The principal contributions of these economists were an addition of a commuting and leisure time constraint in the household utility function (eg. MacDonald, 1979), an interest in polio-centric urban forms (eg. Evans, 1973), and the influence of neighbourhood (eg. Papageorgiou, 1976). However, the fundamental principles still remained the same (Ball, 1985).

The trade-off model was developed under the assumptions of a monocentric city on an isotropic transport plane with a housing market in perfect competition. The basic premise of the model was that house size and access to the city centre were both important determinants of household utility (Muth & Goodman, 1989). Given that transport costs increased from the city centre at a diminishing rate and households always maximise their utility, the model deduced that land prices would fall at a decreasing rate as transport costs rose. Alonso (1964), who pioneered the model, regarded residential location as simply a conflict between spacious living and easy access to the city centre. In other words, how a household balances 'the costs and bother of commuting against the advantages of cheaper land with increasing distance from the center of the city and the satisfaction of more space' (pp. 15). Hence, the optimal location for a household was one where the decrease in housing costs with a move away from the city centre was equal to the increase in transport costs. A stable housing market equilibrium was eventually reached by each household choosing their optimal location through a bid-rent function.

1.4.3.3 The Bid-Rent Function

The bid-rent function was developed by Alonso (1964) and describes how a housing market equilibrium can be derived from the individual location demand functions of the trade-off model. If a household locates by trading off travel costs against housing costs, then associated with this trade off is a particular level of utility which is fixed equal to the maximum utility attainable if the household located at the city centre. To locate away from the city centre, the household needs to be indifferent with the new location if the same level of utility is to be maintained. This is achieved by the household bidding or stating the level of rent per unit housing they are prepared to pay at this new location. Since this price must be low enough to offset transport costs, housing costs decline with distance from the city centre. A bid-rent schedule can be obtained that indicates the relative priorities for rent and travel costs. This bid-rent schedule can be used to calculate a bid-rent curve, which describes bid-rents as a continuous function of distance, whilst holding utility constant. Each household has a complete set of bid-rent curves covering all possible rents and distances. The lower the bid-rent curve, the lower the rent per unit housing and hence the higher the utility. (Muth & Goodman, 1989)

If all the households in the city have the same incomes, tastes and preferences, their set of bid-rent curves will be identical. If the city is in market equilibrium, the rent gradient will lie wholly along one of the bid-rent curves. If it does not, then households would adjust their location in order to maximise their utility until the rent gradient and the bid-rent curve coincide. The particular bid-rent curve that coincides with the rent gradient depends upon residential and non-residential demands for space. However, since households generally have different incomes, tastes and preferences, their set of bid-rent curves will not be identical. In this case, the rent gradient will not lie along a single bid-rent curve, but will be made up of sections of the lowest attainable bid-rent curves of all the households in the city. Hence, a household's optimal location will be the point at which the rent gradient is tangential to their lowest achievable bid-rent curve. At this point, the slope of the rent gradient is equal to the slope of their bid-rent curve and the households utility is at its maximum.

1.4.3.4 Implications and Criticisms of the Trade-off Model

The trade-off model has become a paradigm for much urban economic research. Its great strength lies in its heuristic power in producing results consistent with data from real cities and with findings of earlier urban theory (Evans, 1973; Maclennan, 1982; Muth & Goodman, 1989). In particular, the trade-off model deduces the existence of a negative rent gradient from the city centre outwards, which decreases with increasing distance. It is this feature of the trade-off model that has been the most influential in hedonic house price studies. As will be explained in the next section, one of the motivations behind early hedonic house price research was to estimate this negative rent gradient, as this would strengthen the argument for the concept of the bid-rent function.

Criticisms of the trade-off model are plentiful (Bassett and Short, 1980; Maclennan, 1982). One of the main criticisms is that the trade-off model is demand orientated, with no regard for the supply of housing. With respect to the existing stock, supply has either been ignored or effortlessly adapted to variations in demand, 'almost in the fashion that children build with lego' (Bourne, 1981. pp. 131). Other substantive criticisms concern the fundamental importance of accessibility to the city centre in determining residential location. Other housing attributes, such as housing quality and population density, tend to be broadly correlated with distance from the city centre, and these may have more of an influence over a households choice of location than issues of accessibility. A final criticism is that, whilst neo-classical models are relatively successful at describing residential location patterns, they fail to adequately explain them since they ignore the wider social structures and institutions that govern household decisions. They also suffer from a neglect of the social relations of housing provision in a historically specific context. As Ball (1985) observes, neo-classical models commit 'considerable violence to [our] common-sense understanding of urban spatial structures' (pp. 506).

Section 1.5 Housing Market Disequilibrium and Segmentation.

1.5.1 Introduction

The previous two sections have briefly described some conceptual and economic considerations that have underpinned hedonic house price theory. In particular, the trade-off model of residential location, and the deduction of a negative rent gradient from the city centre outwards, have been important theoretical constructs that have shaped much hedonic house price research. Indeed, the motivation behind the early hedonic house price research was to provide empirical evidence of a negative rent gradient as verification of the trade-off model. However, as will be discussed in subsequent chapters, empirical results generated by early hedonic research were inconsistent and contradictory, particularly in the estimation of the negative rent gradient. Moreover, the assumptions of a housing market in perfect equilibrium, operating under Pareto Optimum conditions, were questioned in early hedonic work. Instead, the housing market was re-formulated in terms of housing market segmentation and disequilibrium, with the concept of housing submarkets becoming important. This section will examine the theoretical and empirical considerations of this view of the housing market. In particular, it will focus on the ideas of imperfect knowledge of buyers and sellers, and the influence of institutions and actors in structuring the housing market.

1.5.2 Housing Submakets

'Heterogeneity in the existing stock, other differences in neighbourhood desirability, and the existence of discrimination imply that the urban housing market is a set of compartmentalised and unique submarkets delineated by housing type and location' (Schnare & Struyk, 1976; pp. 147).

Housing market disequilibrium occurs when changes in demand and supply are unequal. This may occur for a number of reasons. Prospective buyers often have limited house search areas due to search costs, imperfect information, or a desire to be close to workplace, friends or relations. This will mean that only a limited number of housing bundles will be taken into consideration, which can lead to imperfect competition. There may also be highly inelastic demands for certain housing, especially in high quality neighbourhoods, and this could be confounded if the demand is shared by a large number of households. The very nature of housing means that supply is generally inelastic in the short run, and is quite often inelastic for some housing bundles over longer periods due to durability of stock which is difficult to modify, and a lack of building land constraining location. This usually means that housing demand will usually change more rapidly than supply, and this is exacerbated by the time lag in new completions and conversions. The resulting disequilibrium may be pervasive and it will be compounded by investment decisions, causing a high degree of under-occupation and inefficient use of the housing stock (Short, 1982).

Moreover, restrictive supply and demand processes may segment the market into a number of more or less independent sectors, with local supply and demand mechanisms resulting in a different structure of prices in each. These sectors can be viewed in two domains: whether the stock is partitioned into distinct sectors in aspatial terms, or whether the urban area is also geographically subdivided into 'spatial submarkets' (Bourne, 1981). Most commentators now agree that a functional urban housing market does not operate as one large market, but rather as a series of linked, quasi-independent submarkets. Their existence is reflected by significant differences in prices paid for a given amount of housing services. Housing submarkets arise for several reasons. Firstly, they are the result of housing market disequilibrium caused by the factors discussed above. These factors will become exaggerated in larger urban areas through the sheer size and heterogeneity of the housing stock and the diversity of demands placed upon it by a more heterogeneous population. Secondly, they are the result of institutional barriers and are significantly influenced by the actions of gatekeepers such as land-owners, developers, estate agents, housing managers, and financial institutions whose motivation and behaviour largely structure the supply of housing (Knox, 1995). This is particularly important with respect to housing segmentation caused by racial discrimination. The influence of these actors and institutions shall now be briefly examined.

1.5.3 Actors and Institutions in the Housing Market

Supply and demand opportunities are shaped and constrained by various agencies and professional mediators. These have been termed gatekeepers (Saunders, 1990), and represent the institutions and agencies that operate at the interface between the housing stock and buyers and sellers. These include local government agencies, builders and landowners, although the two most documented examples of gatekeepers are mortgage lenders, such as building societies, and exchange professionals, such as estate agents.

Building societies have been documented (eg Boddy, 1980) to have a bias towards certain people, places and types of housing stock when allocating mortgages. In particular, people of colour, those on low incomes or part time employment, and old, large housing in deprived neighbourhoods are less likely to be granted a mortgage. Moreover, this may be translated into a spatial bias, with financial institutions avoiding what they regard as 'risky' areas. This is known as redlining, and is the reluctance to advance funds on any property within neighbourhoods perceived to be a bad risk, usually innercity areas with a high percentage of ethnic minority households and students. However, the affects of redlining are now of dwindling importance in the UK, given the changes in the provision of housing finance during the 1980s (see *Chapter Four*). Estate agents can also influence the allocation and distribution of housing in several ways. Since they control housing market information for both buyers and sellers, they may introduce bias by steering households into or away from specific markets. This is examined in more detail in *Chapter Four*, where estate agents were seen to structure sales and property valuations within specifically defined areas.

Section 1.6 Conclusions

This chapter has introduced the basic aims of the research and discussed some of the underlying themes. It has explained that hedonic house price research developed from the micro-economic theories of location and landuse in the late 1960s and 1970s, and that much hedonic research has been underpinned by concepts of the trade-off model of residential location.. These theories have subsequently been reformulated to take into account the vagaries in the supply and demand of housing, and the influence of actors and institutions upon the housing market. In particular, housing submarkets have become an important concept. Therefore, it can be concluded that in some respects, hedonic house price theory has a better conceptualisation of housing market dynamics than conventional micro-economic theory. This shall now be expanded upon in

Chapter Two and *Chapter Three*. In particular, *Chapter Two* is devoted to the hedonic house price function and the underlying theory and methodology, whilst *Chapter Three* discuss the concept of locational externalities, and how these can be modelled within a GIS.

Chapter Two

The Hedonic House Price Function

Section 2.1 Introduction

This chapter is concerned with the theoretical and conceptual issues of the hedonic house price function. The hedonic house price function relates the price of a house to its attributes via the mechanisms of the housing market. Following certain assumptions, this makes possible the estimation of the implicit price of each housing attribute. However, it is important to note that in recent years, the emphasis of such research has been upon the development and estimation of demand models, as opposed to the continued development of the hedonic house price function. This has resulted in some of the basic problems of the hedonic house price function having been neglected, and moreover, carried through into the work on demand models. Hence, although this chapter will review the main issues concerned with modelling the demand for housing attributes, it will concentrate upon the estimation of the basic hedonic house price function, with particular emphasis upon its of specification.

The chapter is divided into three sections. The first sections deals with the economic theory underpinning the hedonic house price function. The next section is concerned with the incorporation of space into the function, with particular emphasis upon the problems caused by misspecification of the hedonic price function with respect to spatial data. The last section is concerned with methods of contextualising the hedonic house price function using the expansion method.

Section 2.2 Utility Theory

When a commodity is consumed, some benefit or satisfaction is derived. This is called utility. Utility has been conceived as the property of an object that produces benefit, advantage, pleasure and happiness (Veldhuisen & Timmermans, 1984). It is based upon the principle of consumer sovereignty; the belief that individuals are the best judges of their own needs. A consumer will choose a commodity to gain the greatest benefit; to 'maximise his or her utility'. Utility theory identifies a consumer's utility function based on either assumed or revealed preferences and predicts choices constrained by the consumers level of income. Hence, following Freeman (1979a), the conventional utility maximisation problem may be expressed as:

$$\text{maximise } U = U(X) \qquad \textbf{2.1}$$

$$\text{subject to } \Sigma (p_i x_i) = Y$$

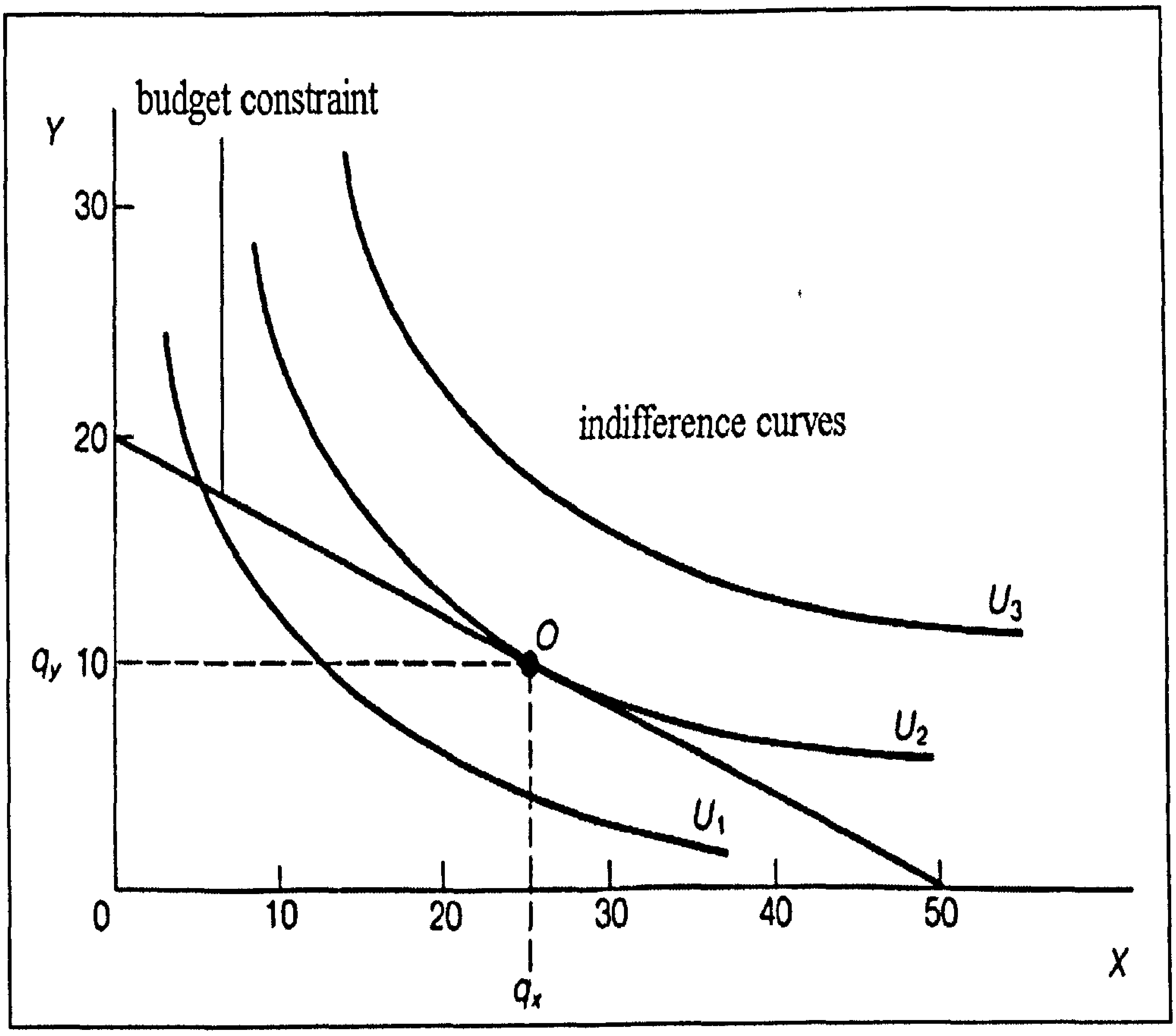
where U is a consumers utility function, X is a vector of commodities ($X = x_1, ..., x_n$), P is a vector of prices ($P = p_1, ..., p_n$), and Y is annual income. The solution to this problem leads to a set of ordinary demand functions conditional on prices and income, and some maximum utility level U_y.

$$x_i = x_i (P, Y) \qquad \textbf{2.2}$$

This is shown graphically in Figure 2.1. Here, an indifference curve (U_2) joins together all the combinations of two commodities (x and y) which yield the same utility to the consumer. The slope of the curve is the marginal rate of substitution, and reveals the combinations of the two commodities to which the consumer is indifferent. To determine which combination is chosen, income levels and the price of the commodities also have to be taken into consideration. This is the budget constraint faced by the consumer, and is shown by the budget line in Figure 2.1. The highest point on the indifference curve that intersects the budget line is called the point of consumer equilibrium, and is the point at which the consumer is maximising his or her utility subject to their budget constraint (point 0 in Figure 2.1)

Figure 2.1

A Hypothetical Utility Function for Two Commodities



Source: Johnston et al., 1994. pp. 277

2.2.2 The 'Characteristics of Goods' Approach

There are many commodities that are not homogeneous but are traded on well-integrated markets. Houses and cars are both good examples. The utility provided by these commodities is based upon the utility yielded by their various attributes, rather than the composite good itself. The theory behind this 'characteristics of goods approach' was developed by Lancaster in 1966, and Griliches in 1971, and later expanded by Rosen (1974) who provided the theoretical framework for analysing a market for a single commodity with many attributes. Following Rosen, differentiated products like houses are assumed to be made up of bundles of attributes, that are not explicitly traded on the market, but as part of a package of housing services. Households are assumed to be utility maximisers and have a strongly separable utility function.

A utility function is strongly separable if it can be partitioned into subsets, and the marginal rate of substitution between two commodities within a subset is independent of the quantities of commodities in other subsets. Consequently, households will divide their incomes between a subset of housing attributes and a subset of non-housing commodities, and will independently maximise their utility for each subset. Hence, if it is assumed that the vector of non-housing commodities can be regarded as one composite commodity, the bundle of housing attributes can be analysed independently. Moreover, the implicit prices of the housing attributes can be revealed by hedonic analysis and these can then be used to estimate the market valuation for particular housing attributes and subsequently the demand for these attributes.

Section 2.3 The Hedonic Price Function

2.3.1 Theory and Overview

Let $Z = (z_1 \ldots z_n)$ be a vector of housing attributes. In Rosen's model of implicit markets, the interaction of supply and demand for Z produces a market clearing function P (Z) which relates the vector of housing attributes to the composite price of the house itself, such that:

$$P(Z) = P(z_1, \ldots, z_n) \qquad \textbf{2.3}$$

P (Z) is the hedonic price function, and describes the house prices resulting from the interplay between housing supply and demand. Buyers and sellers take this price function as given in a competitive housing market. The P (Z) relationship between housing attributes and house prices need not be linear (Harrison & Rubinfeld, 1978). Non-linearities may exist because the housing market may not be in long-run equilibrium since housing supply is generally not very responsive to short term, and indeed long term changes in demand. Moreover, bundles of housing attributes cannot be untied and repackaged to reflect the consumers desired mix of housing services (Rosen, 1974. pp. 37-38).This is an important point. Whilst Lancaster (1966) assumed that the consumer could purchase each commodity in X separately, Rosen argued that it is more reasonable to assume that the suppliers of housing sell bundles of housing services, Z, as part of a package. This has important, and generally unappreciated, implications in the specification of the hedonic price function.

Since the price of a property is a realisation of the price of its housing attributes, P (Z) can be estimated from observations of prices and attribute bundles of different houses. Moreover, the marginal implicit price of any attribute can be found by differentiating the hedonic price function with respect to that attribute. Hence:

$$\delta P(Z) / \delta (z_i) = P(z_i) \qquad 2.4$$

gives the increase in expenditure on Z that is required to obtain a house with one more unit of z_i, *ceteris paribus.* However, it is only under restrictive conditions that the function P (z_i) reflects the household demand for attribute z_i. The estimation of a household's marginal willingness to pay for an additional unit of z_i requires a further stage of analysis, and an understanding of the relationship between the marginal implicit price function, P (z_i), and housing market supply and demand functions.

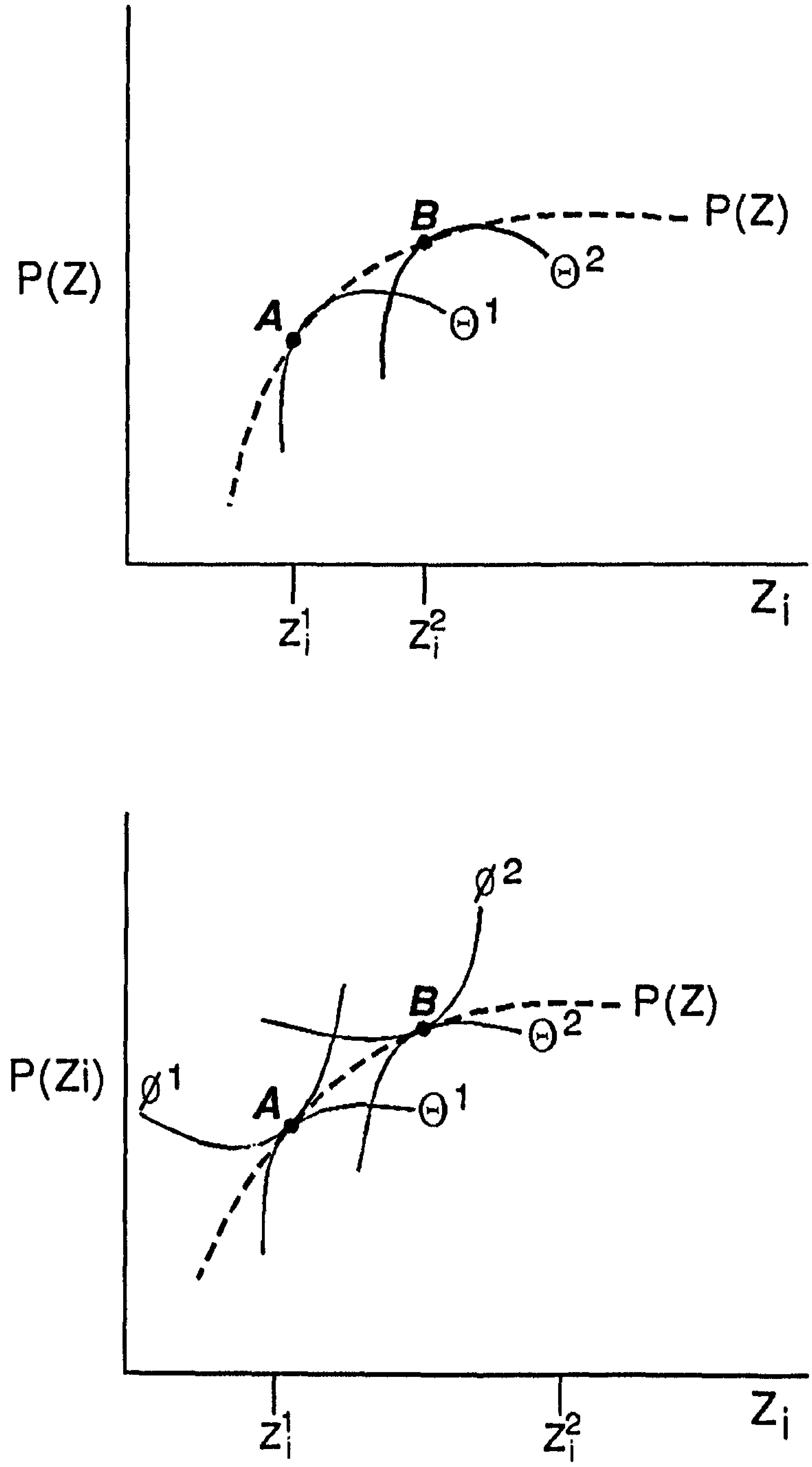
2.3.2 Bid-Rent Functions and Marginal Willingness to Pay

Following Follain and Jimenez (1985), a household is assumed to have a strongly separable utility function U = U(X, Z), where X is the composite commodity of non-housing goods whose price is set equal to one, and Z is the vector of housing attributes. Households then maximise utility subject to the budget constraint Y = P (Z) + X where Y is the annual household income. The partial derivatives of the utility function with respect to a housing attribute is the household's marginal willingness to pay function for that attribute. In other words it represents the additional expenditure a consumer is willing to make on another unit of that attribute and be equally well off.

$$Uz_i / Ux = P(z_i) \equiv \delta P(Z) / \delta (z_i) \qquad 2.5$$

i = 1,...,n, under the usual properties of U

Figure 2.2
Demand and Offer Curves of the Hedonic Price Funtion



Source: Follain and Jimenez, 1985; pp. 79; Fig. 1.

An important part of the Rosen model is the bid-rent function:

$$\theta = \theta\,(z_i, U, Y, \alpha) \qquad \textbf{2.6}$$

where α is a parameter that differs from household to household (i.e. tastes).

This can be characterised as the trade-off a household is willing to make between alternative quantities of a particular attribute at a given income and utility level, whilst remaining indifferent to the overall composition of consumption.

$$U = U\,(Y - \theta, Z, \alpha) \qquad \textbf{2.7}$$

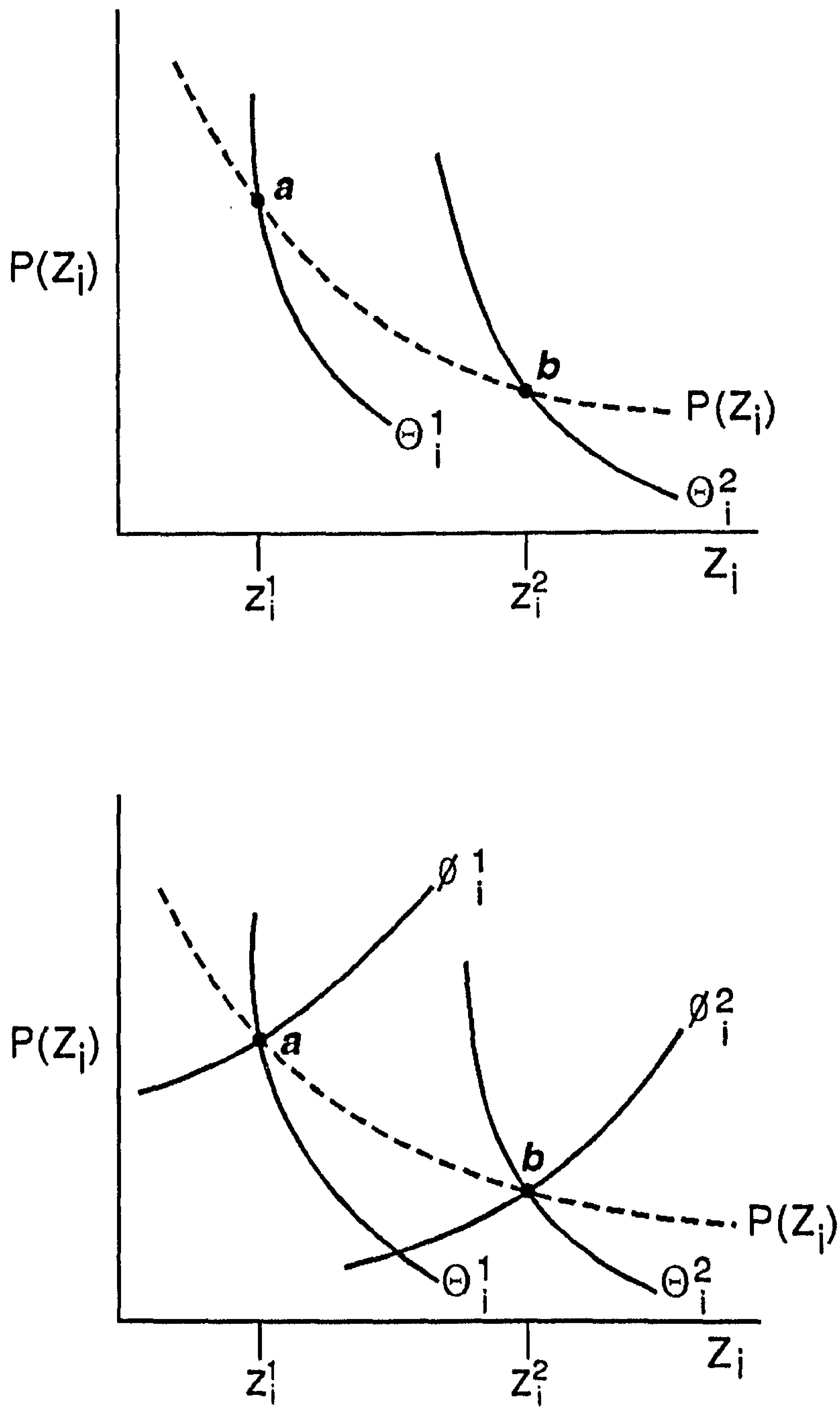
Tracing out these trade-offs generates a household's bid rent function for the given attribute, represented by θ^1 in the upper schedule of Figure 2.2. The household represented by θ^1 is everywhere indifferent along θ^1. θ schedules that are lower correspond to higher utility levels. At maximum utility, the bid-rent curve is tangential to the hedonic price function P (Z). At this point:

$$\theta_i = Uz_i \,/\, Ux \qquad \textbf{2.8}$$

which is the additional expenditure a consumer is willing to make on another unit of z_i and be equally well off (i.e. the demand curve). Figure 2.2 denotes two such equilibria: A for household θ^1 and B for household θ^2

However, since the hedonic price function represents the interplay between supply and demand, the supply side must also be considered. Since, like buyers, suppliers also accept P (Z) as given, then the marginal cost of providing an attribute whilst maximising profits will be a concave offer curve ϕ that is tangential to P (Z). Equilibrium points are those where supply equals demand. Since there are many consumers and many suppliers of housing attributes, there are many bid-rent and offer curves, and so P (Z) represents a function consisting of the joint envelopes of various supply and demand tangencies (Muth & Goodman, 1989). This is shown in the lower schedule of Figure.2.2.

Figure 2.3
The Marginal Implicit Price of an Attribute as a Function of Supply and Demand



Source: Follain & Jimenez. 1985; pp. 79; Fig. 1.

A household maximises utility by simultaneously moving along each marginal price schedule for a vector of housing attributes until it reaches a point where its marginal willingness to pay for an additional unit of each attribute just equals its marginal implicit price (Freeman, 1979b):

$$\theta_i = P(z_i) \qquad \textbf{2.9}$$

This is shown in the upper schedule of Figure 2.3. Hence, if a household is in equilibrium, the marginal implicit prices associated with the chosen housing bundle is equal to the corresponding marginal willingness to pay for those attributes. Thus the marginal implicit price function of an attribute P (z_i), is the locus of marginal supply curves and marginal bid-rent curves for that attribute by different households. This is shown in the lower schedule of Figure 2.3. Rosen's model is very similar to the standard urban (trade-off) model of residential location described in *Chapter One*. This is because the trade-off model can be viewed as a special case of Rosen's model, that focuses on two attributes; access to the city centre and everything else about a house that generates utility.

The hedonic approach assumes that the implicit prices of the estimated hedonic price function, P (z_i) reflects the valuation of attribute z_i as a result of demand and supply interactions of the entire market. However, in general it can be demonstrated that P (z_i) will overstate the inverse demand function for the valuation of an additional unit of the attribute since, since to the right of points (a) and (b), P $(z_i) > \theta_i^1$ and θ_i^2 (see Figure 2.3). Only in extreme cases when all consumers have identical incomes and utility functions will the marginal implicit price curve be identical to the inverse demand function for an attribute. This occurs because P (z_i) is the locus of points on the household's marginal willingness to pay curves θ_i. With identical incomes and utility functions, these points all fall on the same marginal willingness to pay curve (Freeman, 1979b). Hence, the implicit price of an attribute is not strictly equal to the marginal willingness-to-pay, and hence demand for that attribute.

2.3.3 Identifying the Inverse Demand Function for an Attribute.

The above has developed a measure of the price of an attribute as a function of supply and demand interactions in a housing market. But as demonstrated, this does not reveal or identify the inverse demand function for that attribute. The second stage of the hedonic technique is to combine the quantity and implicit price of an attribute to try and identify this function. This is done by expanding the discussion above. Recall that it is only in extreme cases when all households have identical incomes and utility functions that the marginal implicit price function is the same as its inverse demand function. By implication, by taking into account differences in household income, tastes and preferences and other household characteristics which influence utility, it should be possible to adjust the marginal implicit price function of an attribute such that it reflects its inverse demand function. However, two issues need to be addressed. Firstly, identification of the inverse demand function is only possible if the hedonic price function is non-linear with respect to the attribute under investigation. Otherwise the marginal implicit price would be constant and identification of the inverse demand function is not possible (Freeman, 1979b pp. 157). Secondly, the steps necessary to identify an attribute's demand function is dependent upon the assumptions made about the supply side of the market. It is on the latter that the remainder of the discussion will focus.

There are two general possibilities. One approach is to assume that the supply of an attribute is perfectly inelastic with respect to price or willingness to pay. In other words, it is independent of household demand. An inverse demand function could be estimated by regressing equilibrium marginal implicit prices P (z_i) against the quantity of the attribute actually consumed, incomes and other variables (ibid. pp. 165). A second approach assumes that if both the quantities demanded and quantities supplied of an attribute is a function of price, then a simultaneous equation approach can be used to identify the demand function. This is known as the Rosen two-step approach.

2.3.4 The Rosen Two-Step Approach

This is the most popular method in recent literature (Ohsfeldt, 1988), and the one which is closest to Rosen's theoretical model of the implicit market for characteristics. It requires that the marginal implicit price with respect to each attribute, P (z_i), is evaluated for a particular bundle of Z and used as a price vector in a system of demand and supply equations that could be estimated simultaneously. A regression held constant for demand and supply shifts would theoretically yield the inverse demand function for a specified attribute (Follain & Jimenez, 1985). However, there are several problems with this approach. Two of the most important are possible simultaneity bias inherent in the approach, and the identification of the structural parameters (Ohsfeldt, 1988). Efforts to estimate the hedonic price function and to correct these problems has lead to innumerable serious statistical problems (Follain & Jimenez, 1985; Lerman and Kern, 1983; Gross, 1988; Graves et al 1988).

2.3.5 The Bid-Rent Approach

The problem of how to account for supply factors can be overcome if all the households are identical with respect to their income levels and utility functions. It has been previously demonstrated that in this situation, the bid-rent function for an attribute is identical for all households and thus the marginal implicit price function corresponds directly to the inverse demand function for the attribute. Hence, the bid-rent approach involves direct estimation of the bid-rent function rather that first-order conditions associated with a particular bid-rent function (Ellickson, 1981). Basically bid-rent functions are calculated for groups of households which are assumed to receive the same level of utility from the housing bundles being consumed. This assumes that all the households in each group have similar tastes, face the same prices and have similar incomes, thus giving rise to identical bid-rent functions (North and Griffin, 1993). The bid-rent function is estimated by calculating the marginal implicit price curve for an observed attribute. In theory the bid-rent approach is a very convenient and straightforward way to estimate the household demand for housing characteristics. However, its one serious drawback is the difficulty in identifying groups of households with the same utility function (Follain and Jimenez, 1985).

2.3.6 Assumptions of the Hedonic Price Function

There are several assumptions relating to the use of the hedonic price function as a basis for measuring the marginal implicit prices paid by households for bundles of housing attributes (Maclennan, 1982). These assumptions are of necessity very similar to those that underpin the workings of the competitive housing market, as described in *Chapter One,* although the housing market need not function as a unified whole, but maybe in disequilibrium. To summarise:

1. All consumers accurately perceive the attributes represented by the vector Z at every location

2. There is sufficient variation in Z so that the hedonic price function P (Z) is continuous, with continuous first and second partial derivatives

3. Spatial variations in housing attributes are capitalised into differentials in house prices

Of course, any departure from these assumptions may invalidate the supposition that the hedonic price function can be used to estimate a household's valuation of housing attributes

Section 2.4 Hedonic Price Models and the Incorporation of Space

2.4.1 Introduction

The hedonic price model was developed within the framework of economics to study essentially aspatial composite commodities such as cars, refrigerators, washing machines, and personal computers (Griliches, 1971; 1994). However, it has also been extensively used to analyse commodities, such as houses, which have spatial attributes. This has presented a fundamental problem; how to incorporate space into an aspatial econometric model. Neglecting the spatial element of these commodities can result in problems such as spatial heterogeneity, spatial autocorrelation, and multicollinearity which can produce distortions in the econometric model. These problems can also occur if the spatial element is only partially accounted for. This section is concerned with the evaluation of specifications of hedonic models, particularly with respect to their ability to handle spatial data.

2.4.2 The Traditional Specification

Housing attributes are the bundles of housing services that provide utility to the consumer. Fundamentally, Wilkinson (1973b) makes the distinction between dwelling specific or structural attributes and location specific attributes. The former are concerned with factors pertaining to the physical structure of a property, whilst the latter are concerned with the property's location. Hence, the hedonic price function can be defined as:

$$P(Z) = f(S,L) + \varepsilon \qquad 2.10$$

Where P is a vector of observed house prices, S and L are vectors of structural attributes and locational attributes respectively, and ε is a vector of random error terms. Typically, the specification of this function has been defined as:

$$P_i = \alpha X_i + \Sigma \beta_k S_{ki} + \Sigma \gamma_q L_{qi} + \varepsilon_i X_i \qquad 2.11$$

Where:

$i = 1, ..., N$ is the subscript denoting each property;

P_i is the price of property i;

$k = 1, ..., K$ is the number of structural attributes;

$q = 1, ..., Q$ is the number of locational attributes;

α, β, γ and ε are the corresponding parameters;

X_i is a column vector which consists entirely of ones.

This has been termed the traditional hedonic specification (Can, 1992), and has been the basic model in the majority of studies. If the attributes are taken as deviations from their mean, then the model suggests that the price of house i is a function of the average housing market price of a typical property (α), the cost of structural and locational attributes (β_k) and (γ_q), and the price associated with the idiosyncratic elements of the individual house (ε_i). The model is estimated by ordinary least squares (OLS) regression, in which the regression coefficients represent the implicit price of each attribute. Hence, the hedonic price model has to satisfy the following OLS regression assumptions:

1. The relationship between the dependent variable (house price) and the independent variables (housing attributes) are linear in the parameters
2. The independent variables (housing attributes) are free from multicollinearity
3. The errors terms are normally distributed with a mean of zero
4. The error terms are independent, that is, they are not autocorrelated.
5. The error terms have a constant variance; that is, they are homoscedastic.

Any violation of these assumptions can lead to unreliable and biased parameter estimates. As will be seen, the violations of assumptions four and five are common feature of many studies, and are usually caused by the misspecification of the hedonic house price model.

Misspecification issues have tended to be concerned with omitted variables and functional relationships. Hence:

"Finding the correct specification of the hedonic relationship for housing requires that we identify both the correct list of independent variables and the true functional form" (Butler, 1982; pp. 96)

These are important specification issues, since the wrong variables and an incorrect functional form can introduce bias into the model. Debates concerning the range and class of variables are discussed in detail in *Chapter Three*, although at this stage it should be sufficient to note that theory offers little guidance in determining which particular attributes to include in the model (Ohsfeldt, 1988).

2.4.3 Functional Form

A fundamental issue in estimating the hedonic price function is choosing the functional form. A frequent criticism of hedonic studies is that functional form is chosen on the basis of convenience (Halvorsen & Pollakowski, 1982). Unfortunately, theory does not generally suggest a particular functional form for property attributes since, as Rosen has demonstrated, the hedonic price function is a reduced form equation reflecting both supply and demand mechanisms. Hence, the functional form may not be determined from information pertaining to either the underlying supply or demand equations. Instead its shape is determined by the distribution of housing bundle types and household types within a particular market area (Quigley, 1982). However, four functional forms are commonly used in hedonic house price models; linear, semi-log, log-linear and inverse semi-log (Palmquist, 1984).

It is important to impose a functional form which predetermines the correct relationship between the implicit price of a given attribute and the quantity of that attribute. For example, a linear functional form imposes the restriction that the implicit price of an attribute is constant across all quantities of that attribute. So, in the case of the implicit price of energy efficiency improvements (Johnson & Kaserman, 1983), an efficiency improvement in an extremely inefficient house is valued the same as an improvement in an extremely efficient house. Also, the use of a linear functional form requires the implicit price of energy efficiency to be independent of the level of other house attributes such as age and size (Dinan and Miranowski, 1989). More importantly, if the true functional form of the hedonic equation is not linear, the restriction of linearity may result in bias in the resulting coefficients (Linneman, 1980).

In the absence of theoretical guidelines, the Box-Cox, and occasionally the related Box-Tukey generalisation methods of transformations are often specified to search for an appropriate functional form (eg. Freeman, 1979b; Halvorsen & Pollawski, 1981). Freeman (1979b) demonstrated that out of eight alternative hedonic price functions specified, only the Box-Cox transformation allows the implicit price of an attribute to depend upon the level of other attributes and to either decrease or increase as the level of the attribute varies. Although a full Box-Cox model may be specified, in which all the variables may take on a different power transformation factor, the usual procedure is to use a constrained version where all the continuous independent variables have the same power transformation factors, usually within a range of plausible values. However, there is no conceivable behavioural rationale for presuming this 'globally' imposed fit, with the only justification being one of minimising computational expense (Dunn et al, 1987). A full Box-Cox model may be specified as:

$$P(\theta) = \alpha + \Sigma\, \beta_k Z_k^{(\lambda k)} + \mu \qquad \textbf{2.12}$$

Where:

$$P(\theta) = \frac{P^{(\theta)} - 1}{\theta}, \quad \theta \neq 0$$

$$= \ln P, \quad \theta = 0$$

$$Z_k(\lambda k) = \frac{Z_k^{(\lambda k)} - 1}{\lambda_k}, \quad \lambda_k \neq 0$$

$$= \ln Z_k \ , \qquad \lambda_k = 0$$

and Z is the vector of housing attributes.

Then, in the constrained version, all $\lambda_k = \lambda$.

Furthermore, if the values of θ and λ are constrained equal to 1, the model reduces to the linear form. If $\theta$and λ are constrained equal to 0, the model reduces to the log-linear form. If the value of θ is set equal to 0 and the value of λ is set equal to 1, then the semi-log model results. The opposite of the latter specification results in the inverse semi-log. Hence all the restricted functional forms commonly used are subcategories of the Box-Cox model, and statistical tests are used to determine which functional form best suits the data. But there is no reason why the testing of hypotheses should be 'straight-jacketed' into these most common functional forms (Longley & Dunn, 1988). It might also be difficult to choose between two or more specifications which have approximately the same scores on the statistical tests. Moreover, Dunn et al., (1987) have argued that the Box-Cox and Box-Tukey transformations are an undesirably mechanistic means of deriving functional form, and are unnecessarily clumsy and cumbersome in comparison to graphical diagnostic tests and exploratory data-analytic approaches in general. They also note that the Box-Cox and Box-Tukey transformations may not adequately account for the influence of outlying or anomalous data points, although this may be ameliorated by graphical techniques. By demonstrating how partial regression plots and other graphical diagnostics aided in the derivation of the functional form of a logistic regression equation, they were able to conclude that such techniques offered a considerable improvement in flexibility and a greater coherence of interpretability compared to the restrictive Box-Cox and Box-Tukey traditions. However, such interactive data exploration techniques have been lacking in hedonic house price research, despite the introduction of user-friendly computer packages in recent years.

In terms of hedonic house price analysis, where the primary goal is to obtain accurate estimates of marginal prices, the functional form that generates the 'best fit' for the hedonic price function may not be the same as the functional form that generates the 'best' marginal price estimates (Ohlsfeldt, 1988). Indeed, in the work by Halvorsen & Pollakowski (1981), fewer counter-intuitive negative marginal price estimates were obtained using less complex functional forms.

Similarly, in a simulation study by Cropper et al (1988), complex functional forms produced much greater errors in marginal price estimates. They both concluded that a simple linear Box-Cox specification of the hedonic price model generates the smallest errors in marginal price estimates. However, in light of the above discussion on the problems of using such mechanistic and restrictive methods, particularly with respect to unusual data points, such conclusions can be regarded as somewhat naive.

2.4.4 Spatial Misspecifications of the Traditional Hedonic Model

The traditional hedonic specification assumes that the effects of structural attributes on property values are fixed across the housing market, and hence each property will have the same marginal implicit prices. Locational attributes are incorporated as an additional set of housing attributes, independent of the structural attributes. This suggests that a household evaluates the structural attributes of a house, and the attributes of its location, separately. Can (1990) suggests that in this conceptualisation, location can be regarded as an additional premium on the price of a house, independent of the cost of the structural attributes. Furthermore, this specification suggests that there is no interaction or relationship between the structure of a house and its location within a city, which contradicts urban economic theory.

In recent years, it has become apparent that the traditional hedonic specification has not fully captured the spatial element of the data (Can, 1990; 1992). In particular, traditional hedonic models may suffer from spatial dependence and spatial heteroscedasticity. The problems caused by such spatial effects on the validity of traditional statistical methods has long been recognised (Anselin, 1988a). In particular, spatial effects will violate the assumptions of independently, identically distributed errors in the OLS regression model (assumptions Four & Five) used to estimate the hedonic model.

The problem of uncontrolled spatial effects in an hedonic model can be illustrated by the often quoted study of the demand for clean air by Harrisons & Rubinfeld (1978). In this study, the Rosen two-step approach was used to estimate willingness-to-pay curves for air quality improvements. A traditional hedonic specification was estimated. Heteroscedasticity was discovered and the model was subsequently re-estimated using weighted least squares. The model was subsequently analysed by Belsley et al., (1980). Diagnostic tests suggested that spatial autocorrelation was present and that the model may have also suffered from spatial heteroscedasticity. If Belsley et al., (1980) findings are correct, then it can be assumed that Harrisons & Rubinfeld's (1978) willingness-to-pay results are seriously flawed.

Therefore, the two spatial effects, spatial heterogeneity and spatial dependency, must be resolved if the multiple regression models used in evaluating the hedonic house prices are not to be invalidated. However, their effects in mainstream statistical and econometric literature have been almost totally ignored. As Anselin & Griffith (1988) have concluded:

"[E]ven though the methodological results achieved in the fields of spatial statistics and spatial econometrics have been substantial, the dissemination from research community to applied world has been virtually non-existent" (pp. 14)

This is typical of most hedonic house price research, even though the data are likely have inherent spatial structures and be subject to various spill-over effects. This ignorance can be explained in part by the fact that the standard tests for functional misspecification, the selection of variables and the evaluation of predictive performance are not affected by the spatial nature of the models and data (Anselin, 1988a. pp. 282).

2.4.5 Spatial Heterogeneity and Housing Submarkets

"*The central problem in estimating hedonic equations involves the delineation of homogeneous submarkets*" (Straszheim, 1974; pp. 404).

The assumption that structural attributes will have the same fixed marginal implicit prices across urban space implies the presence of a single homogeneous competitive market. This fails to take into account the housing market dynamics that can lead to submarket formation. Generally, property prices have been conceived as varying continuously across urban space. Warnings against such a view, such as by Schnare & Struyk (1976), have generally been ignored. However, urban space is divided up into discrete units by transportation routes, housing stock and landuses. Spatial spill-over effects from these units implies that property prices are better conceived as contiguous rather than continuous. Since, by definition (see *Chapter One*), each of the submarkets will have a unique supply and demand structure, the implicit prices of the attributes will no longer be constant, but vary by submarket. If uncontrolled for, this spatial heterogeneity of implicit prices will cause structural instability in the regression coefficient and error term (Can, 1992). The result is a special case of heteroscedasticity which will violate the assumption of constant error variance.

Previous studies (e.g. Ball & Kirwin, 1977; Schnare & Struyk, 1976; Goodman, 1981) have tried to deal with spatial heterogeneity caused by the presence of submarkets by the method of 'switching regression' (Can, 1992). This involves estimating an hedonic house price model for the entire housing market and then separate ones for each submarket, with the specification of the model only concerned with the structural attributes of each house. Hence, if the housing market is divided up into 'M' discrete submarkets, then:

$$P_j(Z) = f_j(S) + \varepsilon_j \qquad j = 1, \dots, M \qquad \textbf{2.13}$$

Where P_j is the vector of house prices in submarket j
S is the vector of structural attributes.

If a statistically significant difference exists between the estimated coefficients for the entire market model the coefficients in each of the submarket models, then this may provide evidence for structural instability and thus indicate the existence of a fragmented housing market.

Unfortunately, the method is highly complex and somewhat arbitrary. There are two main problems: identification and verification of potential submarkets. The method requires submarkets to be identified *a priori*, but this is problematic due to the hidden nature of the geography of supply and demand mechanisms that determine the implicit prices of the attributes. An indication of differential submarket mechanisms can be gained from an examination of house price inflation across the housing market. If there are varying demand and supply processes at work, then it can be expected that relative house price inflation will vary and houses in some areas will increase in price faster than others (Munro & Maclennan, 1987). In practice however, submarkets have generally been identified by either housing attributes or along neighbourhood definitions. The former has involved segmenting the dataset by the attributes of the housing stock, such as property type or tenure, whilst the latter by defined areas, typically based upon existing geographies such as political boundaries or census areas. A further stratification scheme has been to segment the data set by household characteristics such as race or socio-economic class, since these will influence the geography of the supply and demand schedules that are being defined. Schnare and Struyk (1976) suggest experimenting with several stratification schemes, using any large and significant differences in estimated parameters as evidence of submarket existence. Goodman (1981) suggests a more objective approach, using a method of defining submarkets based upon choosing the fewest number of submarkets possible, with the housing bundles within each being as similar as possible. He also suggests that since public services are an important part of a neighbourhood, houses within a given municipality should be placed in the same submarket. He argues that this should minimise the effect of differences in the cost and supply of public services between each municipality. Unlike Schnare and Struyk, Goodman (1981) suggests using analysis of covariance as verification of the submarket. He also points out that, since submarkets reflect differences in supply and demand schedules across the housing market as whole, the functional form of the hedonic model may also vary between submarkets.

The influence of submarkets has been contradictory. Whilst Goodman (1979; 1981) concluded that the implicit prices of attributes vary between submarkets, Schnare & Struyk (1976) and Ball & Kirwin (1977) discovered that the differences between the estimated coefficients for the submarket models and the model estimated for the whole housing market were insignificant. However, as discussed in *Chapter One*, urban housing markets are unique, and hence submarkets may form in some urban areas but not others, and this may explain the contradictory results. Furthermore, the arbitrary nature of defining submarkets by administrative geographies may lead to significant heterogeneity of supply and demand schedules within so called homogeneous submarkets. Also, the disaggregation of the housing market into discrete areas may be unrealistic, since the influence of certain locational attributes will extend beyond submarket boundaries. This is discussed in more detail in *Chapter Three*.

2.4.6 Spatial Dependence

The second problem is one of spatial dependency or spatial autocorrelation. Spatial autocorrelation will violate the assumption of independent errors. This could lead to misleading inferences about the significance of parameter estimates and can also negatively effect the validity of a wide range of standard diagnostics test. It occurs in regression analysis by two distinct forms of misspecification. Firstly, a variety of common misspecifications can result in spatial autocorrelation. Anselin (1988b) summaries these as factors associated with spatial aggregation, the presence of uncontrolled for non-linear relationships, and the omission of relevant variables. An example of the first is when aggregation of data results in spatial heterogeneity, such as treating the housing market as a unified whole instead of as a series of submarkets, and this produces spatial autocorrelation in the error term.

The second cause of spatial autocorrelation results when spatial data is incorrectly modelled. This is more fundamental in the sense that it is a special feature of spatial data (Can, 1990). It occurs in hedonic research since, firstly, the prices of nearby houses are similar because they share common locational attributes and will tend to have similar structural attributes, and secondly, because the prices of nearby houses will have an absolute or externality effect upon each other. The first consideration is the basis of house price theory, and will only be problematic with respect to omitted locational attributes. The second is less tangible. Can argued that the workings of the housing market were such that estate agents and buyers would base the price of a house not only upon its structural and locational attributes, but also on the prices of properties in the immediate vicinity. In the same way, home owners may forego certain home improvements if they perceive that the affect on the capital value of the property will be minimal with respect to house prices in the immediate area. This was corroborated by anecdotal evidence from estate agents in Cardiff[1] who expressed the problems of selling housing that had been 'upgraded beyond the selling price of the area that it was located in'. Can described these as 'adjacency effects'.

[1] See *Chapter Four* for details of the estate agent survey

A third complicating factor in model specification is the joint occurrence of both spatial effects, since factors which cause spatial autocorrelation are also likely to lead to spatial heterogeneity (Anselin, 1988a. pp. 290)

Section 2.5 Contextual Hedonic Models

2.5.1 Introduction

The effects of spatial data in hedonic models have been generally ignored. This is despite the fact that such effects are pervasive and can be expected to be a fundamental feature of house price data. The traditional hedonic specification attempts to capture spatial variation solely by the use of locational attributes, which historically have been poorly specified - see *Chapter Three*. But the factors that contribute to spatial heterogeneity and spatial dependence are inherent in the structure of the data and the mechanisms of the housing market. Thus, Can (1990; 1992) has argued that the spatial aspects of house price data should be modelled explicitly within the specification. This requires an alternative specification of the hedonic function, that will have to take into account both submarkets (spatial heterogeneity) and the price of adjacent houses (spatial dependence). Such specifications can be regarded as 'contextual hedonic models' since they take into account the context of the housing market. Two types of contextual models have been proposed: spatial expansion hedonic models (Can, 1990;1992) and multi-level hedonic models (Jones & Bullen, 1993;1994).

2.5.2 The Spatial Expansion Specification

Can developed a series of hedonic model specifications to deal with spatial effects that were based upon the expansion method (Casetti, 1972. 1992). The expansion method is a technique for generating mathematical and statistical models by expanding the parameters of more simpler models. It has become a well known procedure for 'contextualising' existing models (Foster, 1991), and has been used 'to ascertain how, where, when and why [functional] relationships vary from context to context' (Jones, 1991. pp. 45). Since this can include spatial context, Odland et al (1989) characterised the expansion method as a way to address specification errors arising from spatially heterogeneous processes, such as housing market dynamics. The expansion method inherently allows for the presence of spatial heterogeneity (submarkets), and can eliminate the part of heterogeneity resulting from spatial structural instability.

2.5.2.1 Parameter Drift

An important concept of the expansion method is 'parameter drift'. Parameters are said to 'drift' if their estimates significantly differ with context. Thus, parameter estimates of housing attributes can be thought to drift across submarkets. In the traditional hedonic specification, these parameters are assumed to be stable and invariant. '[The] bias is towards presupposing parameter stability, while the opposite should be true' (Cassetti, 1992: pp. 35). Furthermore, Cassetti argues that 'in the social sciences, functional relationships are likely to represent subsystems that will perform differently in different environments and circumstances rather than invariant laws' (ibid.). This suggests that spatial context can not be simply equated or controlled for through *ceteris paribus* conditionality. Interaction between the estimated parameters of a dwelling's structural attributes and its spatial location is needed if the contextual 'real world' is to be captured.

2.5.2.2 Using the Expansion Method to Incorporate Space

Foster (1991) suggests that since any location in time and space is unique, reference to a location will provide a unique reference to a given context. 'Spatial' drift will only occur if the contextual variables are themselves systematically distributed through space. Otherwise, 'contextual' drift is said to occur. Cassetti (1992) identified a three stage procedure for generating expanded models. In the first stage, an initial (global) model is generated, in this case the traditional hedonic model. Results from this model can be regarded as characteristic of the average context which is uniformly applied to all observations. The second stage involves expanding some or all of the parameters in the initial model by equations that redefine them as functions of other variables. So in the case of the hedonic model, the structural attributes may be redefined as functions of locational attributes. These are called expansion equations. In the final stage, a terminal model is generated by placing the expansion parameters into the initial model. This is called an expanded model, or a spatially expanded model if the expansion equations contain a spatial element. In this way, space has been explicitly incorporated into the specification.

2.5.2.3 Expanding the Hedonic Price Function

Can (1990) argued that the structural parameters may take different values across urban space, varying with respect to location. She cites the example of a two car garage, which in the traditional specification will have the same marginal implicit price in an inner city neighbourhood, characterised by low car ownership, as in a suburb where the demand will be greater. Thus, using the three stage method, the hedonic price function may be expanded using the traditional hedonic specification as the initial model and location as the expansion equations. This is still problematic though, since the functional form and the attributes to be included in the expansion equations cannot be known *a priori.* Thus:

$$P(Z) = f(L(f(S))) + \varepsilon \qquad \textbf{2.14}$$

where L is a measure of location and S are the structural attributes

This is the spatial expansion hedonic function.

Since the spatial expansion method can be used to capture the spatial heterogeneity of submarkets, a typical expansion specification is one that accommodates discrete space. For instance, the intercept in the traditional hedonic model (equation 2.11) may be allowed to vary for M submarkets:

$$\alpha = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_{m-1} D_{m-1}$$

$$= \alpha_0 + \Sigma\, \alpha_j D_j \qquad \textbf{2.15}$$

where j = 1, ..., M-1

The intercept term has been expanded with respect to a series of (M - 1) dummy variables (D_j) that represent the individual submarkets. Each of the intercept terms ($\alpha_1 \ldots \alpha_{m-1}$) represent the submarket differentials from the base submarket price of α_0. If the expansion equation is placed into the initial equation (equation 2.11), the spatially expanded hedonic specification is as thus:

$$P_i = \Sigma\,(\alpha_0 + \alpha_j D_j)\,X_i + \Sigma\,\beta_k\,S_{ki} + \Sigma\gamma_q\,L_{qi} + \varepsilon_i\,X_i \qquad \textbf{2.16}$$

The average house price varies between submarkets, but since the structural and locational attributes have not been expanded, the functional relationship between price and the housing attributes are invariant across space. To correct for this, the model may be re-specified such that the structural attributes, say, may also vary across submarkets.

$$\beta_k = \beta_{10} + \beta_{11}D_1 + \beta_{12}D_2 + \ldots + \beta_{20} + \beta_{21}D_1 + \beta_{22}D_2 + \ldots + \beta_{Kj-1}D_{j-1} + \beta_{Kj}D_j$$

$$= \beta_{k0} + \Sigma\,\beta_{kj}D_j \qquad \textbf{2.17}$$

where: k = 1, .. K, and j=1,..,M-1

Substituting this into equation 2.16:

$$P_i = \Sigma\,(\alpha_0 + \alpha_j D_j)\,X_i + \Sigma\,(\beta_{k0} + \beta_{kj}D_j)\,S_{ki} + \Sigma\gamma_q\,L_{qi} + \varepsilon_i\,X_i \qquad \textbf{2.18}$$

Not only does average house price vary between submarkets, but so do the marginal implicit prices of the structural attributes. The locational attributes can be expanded in a similar way.

$$P_i = \Sigma\,(\alpha_0 + \alpha_j D_j)\,X_i + \Sigma\,(\beta_{k0} + \beta_{kj}D_j)\,S_{ki} + \Sigma\,(\gamma_{q0} + \gamma_{qj}D_j)\,L_{qi} + \varepsilon_i\,X_i \qquad \textbf{2.19}$$

Jones & Bullen (1994) have pointed out that the result is equivalent to fitting a separate regression model between price and housing attributes for each submarket, which is very similar to the switching regression approach. Moreover, they stress that these two expansion models (equations 2.16 & 2.19) are nothing other than *ANOVA* and *ANCOVA* models. The major conceptual difference between the fully expanded model (2.19) and the switching regression model (2.13) is that the former has a single error term and thus assumes that the error variance is constant throughout the housing market. Since switching regression estimates separate models for each submarket, this assumption is relaxed. If this is the case, then the fully expanded model will be more efficient than the set of separate regressions. However, as Jones & Bullen (1994) argue, there may be no real gain in stringing together the separate regressions in the fully expanded model since the estimated coefficients will be identical, although the estimated standard errors of the coefficients may differ if the latter model is indeed more efficient.

The discrete spatial expansion methods above indicate that implicit prices vary with submarkets. An alternative expansion is to contextualise discrete space with locational attributes. Can (1992) operationalized this by constructing a measure of neighbourhood quality from census data and used this as a locational attribute. Two expansion equations were then specified, based upon a linear and quadratic functional form. The linear expansion equation of β can be specified as thus:

$$\beta_k = \beta_{10} + \beta_{11}NQ + \ldots + \beta_{20} + \beta_{21}NQ + \ldots + \beta_{K0}NQ + \beta_{K1}NQ$$

$$= \Sigma\,(\beta_{k0} + \beta_{k1}NQ) \qquad \textbf{2.20}$$

where $k = 1, \ldots K$ and NQ is a measure of neighbourhood quality

Hence, placing the expansion equation into the initial model:

$$P_i = \alpha\, X_i + \Sigma\,(\beta_{k0} + \beta_{k1}NQ)\, S_{ki} + \varepsilon_i\, X_i \qquad \textbf{2.21}$$

The terminal model (equation 2.21) asserts that location has no implicit price, but instead is seen as 'driving the spatial variation in the housing price determination process' (Can, 1990. pp. 258). Such a concept is supported by Witte et al, (1979), who have argued that implicit markets do not exist for locational attributes, since they are not determined by the action of any single supplier, but rather are the result of multiple independent decisions of the inhabitants. Instead, they serve to shift both the supply and demand curves of structural attributes, and hence the implicit price of the attributes. However, this is debatable, and it will be argued that the implicit markets do exist for locational attributes, and thus implicit prices, but not at the level of the individual house.

The specification incorporates the effects of spatial heterogeneity by allowing the structural attributes to 'drift' with neighbourhood quality, as measured by census data. However, this implies that the geography of the census tracts is a good proxy for submarkets, and moreover, that the spatial heterogeneity of submarkets can be adequately captured by neighbourhood quality differentials, and that the relationship between parameter drift and neighbourhood quality is consistent across the housing market. Also, since the intercept is left unexpanded, the model implies an average house price for the entire market, independent of neighbourhood context and submarkets. As such, Can (1992) concluded that although the spatial expansion hedonic specification reduced heteroscedasticity in comparison to the traditional hedonic specification, it was not completely removed

Spatial dependence caused by the spill over effects of nearby houses cannot be ameliorated by the expansion method alone. Instead, Can (1992) proposed incorporating an autoregressive function into the spatial expansion specification.

$$P_i = \alpha X_i + \rho\, WP + \Sigma\, (\beta_{k0} + \beta_{k1} NQ)\, S_{ki} + \varepsilon_i X_i \qquad \textbf{2.22}$$

Where W is the generalised weight matrix; WP is the spatially lagged dependent variable (house price), and ρ is its coefficient.

The hypothesised spatial dependence is determined by W which is specified *a priori*. In this specification the price of a house is dependent upon the price of properties at nearby locations in addition to its structural and neighbourhood attributes. The coefficient ρ measures this absolute price effect of nearby properties. Since the specification contains a lagged dependent variable WP, OLS regression cannot be used, since the assumption of independent errors will be violated. In this case, OLS would lead to both biased and inconsistent parameter estimates. Can (1992) proposed using a maximum likelihood estimator, and concluded that hedonic models that contained a lagged dependent variable no longer suffered from spatial dependence.

2.5.3 The Multi-level Specification

2.5.3.1 Introduction

The spatial expansion method incorporates contextuality into the specification by expanding what Jones (1991) has described as the 'fixed' parts of the hedonic model. These refer to the parameters of the attributes and the intercept term, which are 'fixed' or unchanging as opposed to the error term, or random part of the model which is taken from a distribution. Jones & Bullen have argued that fixed-part expansions are problematic:

'[Neighbourhood] quality has been made an attribute of each housing unit, with no distinction between houses and the [submarkets] in which they are located. [Submarkets] and houses are treated as equivalent observations, although houses are likely to be more numerous than [submarkets], and houses within a [submarket] are likely to be more similar than houses in a different [submarket]. When there is only one observation per [submarket], the within-place variation is totally confounded with the between-place variation and no separate estimates of these distinct components is possible.' (Jones & Bullen, 1994. pp. 255).

The argument is concerned with the differentiation between compositional and contextual effects (Jones & Bullen, 1993). The major problem is that the contextual effects, that is the difference a place makes, are potentially confounded with the compositional effects, or the differences produced by the variations in housing attributes within each place. Such was acknowledged early on by Wilkinson (1973a), who argued that "[i]t is difficult conceptually as well as statistically to distinguish the effects of the characteristics of a dwelling alone on price since an obvious and important feature of a neighbourhood is its stock of dwellings" (pp. 76). This has been the result of treating the location of a property in one dimension or spatial level. Instead, Jones & Bullen (1993) conclude that, in this case, there are two distinct levels of analysis: houses (level 1) and submarkets (level 2), whereas the spatial expansion methods presume only one single level.

'The single level model assumes that the data does not have a hierarchical structure, that all the relevant variation is at one scale, that there is no auto-correlation and that there is a single general relationship across space and time ... [T]his model denies geography and history; everywhere and anytime is basically the same!' (Jones, 1991. pp. 8)

Inferential errors are likely to occur when inappropriate single-level models are used, and when multi-level data are modelled using techniques designed for a random sample, such as OLS regression. These problems can be overcome by specifying the model, not as varying at a single level, but as varying simultaneously over a number of levels. This is achieved by a modification of the three-stage expansion method. If the error term, or random part of the model is expanded as opposed to the fixed-part, then the expansion can be specified at a higher level. The results of such an expansion is known as a multi-level model (Goldstein, 1987).

2.5.3.2 Using the Multi-level Method to Incorporate Space

After Jones & Bullen (1994), the traditional hedonic specification can now be written in terms of fixed and random parts, algebraically detailing the house and submarket levels. For the sake of clarity, locational and structural attributes have been reduced to a vector Z of K housing attributes.

$$P_{ij} = \alpha_j X_{ij} + \Sigma \beta_k Z_{kij} + \varepsilon_{ij} X_{ij} \qquad \textbf{2.23}$$

Where: $j = 1, \ldots, M$

$i = 1, \ldots, N$

$k = 1, \ldots, K$

where Z is the vector of K housing attributes, and M is the number of submarkets, and N is the number of properties. This is a micro-model, since it is based upon individual data. To achieve the equivalent multi-level model to that of the discrete-space fixed expansion of equation 2.16, the intercept has to be allowed to vary in a higher, level 2 between submarket, model, by the expansion equation:

$$\alpha_j = \alpha + \mu_{\alpha j} \qquad \textbf{2.24}$$

This is a macro-model, since it is based upon aggregated data. The price of the typical house in submarket j (α_j), is seen as a function of the market-wide price, α, plus a differential for each submarket, $\mu_{\alpha j}$. The micro-model is the within-place equation, whilst the macro-model is the between -place equation in which one of the parameters of the within-place model, in the case the intercept, is the dependent variable. Both models combines to form the terminal model:

$$P_{ij} = \alpha_j X_{ij} + \Sigma \beta_k Z_{kij} + (\mu_{\alpha j} X_{ij} + \varepsilon_{ij} X_{ij}) \qquad \textbf{2.25}$$

The two random terms in the brackets are assumed to be independent of each other (Goldstein, 1987). Since the intercepts are allowed to vary according to a distribution, Jones (1991) termed these random intercepts models. Fully-random multi-level hedonic models also allow the attribute parameters to vary according to a higher level distribution. This is achieved by specifying an additional macro-model :

$$\beta_{kj} = \beta_k + \mu_{\beta kj} \qquad \textbf{2.26}$$

This conceives the attribute's implicit price as an average market-wide price plus a submarket differential. The combination of the initial model and the two macro-models produces the terminal model:

$$P_{ij} = \alpha_j X_{ij} + \Sigma \beta_k Z_{kij} + (\mu_{\alpha j} X_{ij} + \mu_{\beta kj} Z_{kij} + \varepsilon_{ij} X_{ij}) \qquad \textbf{2.27}$$

The model now has four random terms; σ^2_{ε} at level 1; and $\sigma^2_{\mu\alpha}$, $\sigma^2_{\mu\beta}$ and the covariance term, $\sigma_{\mu\alpha\mu\beta}$ at level 2. The covariance term allows the random intercepts and attribute parameters to co-vary according to a higher level, joint distribution.

It is this concept of a higher level distribution which is the key to multi-level models (Goldstein, 1987). The differential intercepts and implicit prices are not specified as fixed, separate and independent as in the usual fixed-part expansion model, but as coming from a distribution at a higher level. Since these distributions concern not houses but submarkets, it is identical to treating places as a sample drawn from a population. The means of these higher level distributions are simply the usual intercept and implicit prices representing the average market-wide relationship. It is the variances / covariance's of the higher-level random terms that capture parameter drift. Moreover, if these variance terms are effectively zero, there is no parameter drift and there is no need for macro-models. There are no significant submarket effects and the traditional hedonic specification is adequate in describing house price variation.

2.5.3.3 Multi-level Models and Spatial Effects

Since the model contains more than one error term, it cannot be estimated using OLS regression. Instead, a procedure using an Iterative Generalised Least-Squares (IGLS) algorithm can be used (Goldstein, 1987; Jones, 1991). This algorithm simultaneously estimates the fixed and random parameters in a sequence of linear regressions until it reaches a convergence. An advantage of IGLS is that, unlike OLS, IGLS explicitly models spatial dependence and spatial heterogeneity.

With multi-level data, such as houses nested in submarkets, spatial dependence can be treated as the norm since individual houses in the same submarket are likely to be more similar, in some way, than houses drawn from the entire housing market at random. Hence, autocorrelation is to be expected in hierarchical data, and the multi-level approach exploits this dependence to derive improved estimates, while the standard errors of the estimates are adjusted to take into account the autocorrelation (Goldstein, 1987). The degree of autocorrelation in multi-level models can loosely be conceived as the ratio of 'variation at the higher level' to the 'total variation of all levels'. If the ratio is zero, there is no autocorrelation and only a single level model is needed.

With respect to spatial heterogeneity, multi-level modelling can be seen as a technique that models the structure of the variation that is not accounted for by the housing attributes, and does not assume a constant variance that can be captured by a single error term like OLS regression. Jones & Bullen (1994) illustrate this by demonstrating that the total variance for equation 2.27 between the submarkets at level 2 is the sum of the two random variables:

$$\mathrm{Var}\,(\mu_{\alpha_j} X_{ij} + \mu_{\beta_{kj}} Z_{kij}) = \sigma^2_{\mu\alpha} X_{ij} + 2\sigma_{\mu\alpha\mu\beta} Z_{kij} + \sigma^2_{\mu\beta_k} X_{ij}^2 \qquad \textbf{2.28}$$

The total variance is not constant but a quadratic function of the attributes. In such a specification, the greatest variation between submarkets may be for large and small properties, say, with the smallest spatial variation in prices for properties of average size.

Furthermore, multi-level modelling is capable of capturing heteroscedasticity caused by more common mis-specifications, such as heteroscedasticity caused by an attribute that has been draw from a population with a non-constant variance. For instance, in the case where the relationship between price and size is more variable for larger properties, regardless of submarket location. Traditionally, this has been tackled by transformation, such as in weighted least squares regression (e.g. Harrison & Rubinfeld (1978)), but this is not modelling the heteroscedasticity directly within the specification. Multi-level modelling can overcome this by expanding the random part of the micro-model at level 1 (houses), to give a multi-level model with an additional random term for property size (Jones & Bullen, 1994):

$$P_{ij} = \alpha_j X_{ij} + \beta_0 Z_{0ij} + \Sigma\, \beta_k Z_{kij} + (\, \mu_{\alpha j} X_{ij} + \mu_{\beta kj} Z_{kij} + \varepsilon_{\beta 0ij} Z_{0ij} + \varepsilon_{ij} X_{ij}\,) \qquad 2.29$$

Hence, the attribute for property size, Z_0, is also included in the random part of the model.

2.5.3.4 Other Features of the Multi-level Hedonic Model

Since the random-part differentials can be seen as coming from a distribution, it has certain practical benefits that are lacking in the traditional and spatial expansion specification estimated by OLS regression. Firstly, unlike the spatial expansion specification, which in effect fits a different regression model for each submarket, the multi-level specification uses all the data in the estimation. This means that the implicit prices of the attributes and the intercept term are based on information from all the submarkets. Secondly, this pooling of information results in 'borrowing of strength', whereby submarket-specific relations which are poorly estimated on their own benefit from information from other submarkets. Thus thirdly, this will result in precision-weighted estimation, whereby unreliable submarket specific implicit prices are differentially shrunk towards the overall market-wide estimate. A reliably estimated within submarket relation will not be affected by this shrinkage. Hence, these three features yield substantial benefits in implicit price estimation. Moreover, the pooling of information and borrowing of strength is more coterminous with the definition of submarkets as being quasi-independent and generating externality effects, more so than the switching regression or spatial expansion specifications, that in effect treat them as separate and autonomous. Hence, the estimation procedure is both specific and general, allowing an appropriate compromise between specific estimates for different places and the overall fixed estimate that pools information across places over the entire sample.

2.5.3.5 Extending the Multi-level Hedonic Specification

The multi-level specification may be elaborated upon in several ways. Firstly, the number of levels may be expanded to include different housing markets or different time periods. Secondly, additional attributes can be included in the micro-model, such as structural attributes, or in higher level macro-models, such as locational attributes. For instance, with respect to Can's (1990; 1992) work , if house prices are perceived to drift with neighbourhood quality, then this quality measure can be specified at the higher, submarket level. Hence:

$$\alpha_j = \alpha + \gamma\, NQ_j + \mu_{\alpha j} \qquad \textbf{2.30}$$

In this random intercepts expansion, submarket average prices vary with neighbourhood quality.

2.5.3.6 A Comparison of the Contextual Hedonic Models

Jones & Bullen (1994) compared the spatially expanded hedonic specifications with the multi-level specifications using house price data for London, divided into several districts (submarkets). A comparison of the estimates of the implicit attribute prices revealed that the multi-level specification produced more robust results. In particular, the multi-level model estimated coefficients gained from precision weighted shrinkage in submarkets that had small sample sizes, compared to the spatially expanded model coefficients estimated using OLS which 'does not differentiate whether two or two thousand houses are involved' (pp. 261). In this case, the coefficients had unreasonably high estimates. Moreover, several slope coefficients for house size in the spatially expanded specification were negative, (suggesting paying less for a larger property), and these were shrunk to positive multi-level estimates in the majority of cases. Jones & Bullen (1994) demonstrated that the shrinkage was controlled by both the submarket sample size and the distance of the fixed-part estimate from the overall multi-level (London) average. It was also demonstrated that, unlike the spatially expanded counterpart, the multi-level specification was better able to deal with heteroscedasticity and autocorrelation by modelling it directly, and it was concluded that individual prices were indeed more variable for larger properties. Such variation could not be captured in the single parameter error term of the spatially expanded specification.

2.5.3.7 The Geographically Weighted Specification

Before the chapter is concluded, it is worth mentioning a further specification of the hedonic house price function which explicitly incorporates space. This is the geographically weighted hedonic specification, and is based upon the recent concept of geographically weighted regression (Brunsdon et al, 1996). Similar to the spatial expansion method and multi-level modelling, geographically weighted regression assumes that the parameter estimates drift across space. It then models this variation using an OLS regression model, which is weighted according to the location of each observation in space. The weighting system is calibrated using the proximity of each observation to all other observations within a specified distance, which can vary depending upon locational context. This allows local relationships to be estimated. However, the technique is in its infancy, and is still under development. Therefore, this research will not be using the geographically weighted hedonic specifications, although as is discussed in *Chapter Nine*, this has the potential for future research opportunities.

Section 2.6 Conclusions

This chapter has reviewed the economic theory underpinning hedonic house price models, and has put the spatial element of such models into context. Specifically, it has discussed the problems encountered when using spatial data. Such problems have been generally ignored, and most recent research has concentrated upon using the parameters estimated by the hedonic house price function as the basis for demand equations in an attempt to estimate the demand for housing attributes. However, if the hedonic model suffers from spatial effects, the estimated parameters may be incorrect and hence, so will the subsequent demand equations.

To ameliorate these problems, alternative 'contextual' specifications were developed with the aim of modelling the spatial structures in the data explicitly. These were developed by expanding the fixed and random terms of the traditional hedonic specification to create the spatially expanded specification and the multi-level specification respectively. Although models using these specifications were advantageous over the traditional specification, it was concluded that the multi-level specification was conceptually and empirically more proficient when it came to modelling he spatial structures of the housing market. The next chapter will continue the theme of spatial data by examining the variable specification of the hedonic model. In particular, it will examine the problems associated with the measurement of locational attributes, and how Geographical Information Systems can overcome some of the issues of spatial resolution.

Chapter Three
Housing Attributes and Spatial Data

Section 3.1 Introduction

This chapter is concerned with the variable specification of the hedonic model, and some of the problems encountered with modelling such data. In *Chapter Two*, it was argued that housing attributes could be divided into structural attributes that pertained to the physical qualities of a house and locational attributes that are related to its location. The price of a property is then a realisation of the price of these attributes, although they are never individually identified in property transactions. However, it will be argued in this chapter that the measurement of these variables often fails to take into account the complexity of effects influencing house price, especially with respect to locational attributes, and underestimates the problems associated with the measurement of spatial data in general. The chapter is divided into four main sections. The first section deals with the types of attributes that affect house prices, with particular attention towards the concept of locational externalities. The second section deals with the measurement of housing attributes, within the context of previous research, with the emphasis upon the inconsistencies with regards locational attributes. The third section discusses the problems encountered with modelling spatial data using the hedonic price function, and how this has been historically limited by the resolution of the data. Finally, the last section discusses how Geographic Information Systems may overcome some of the problems associated with the measurement of locational externalities and the use of spatial data in general.

Section 3.2 Housing Attributes

3.2.1 Introduction

Housing attributes have been traditionally divided into structural attributes and locational attributes. However, as will be seen, structural attributes have been far easier to account for in the price of the house than locational ones. Hence, a significant part of this section is devoted to locational attributes since, as will become apparent, they are far more difficult to reconcile than their structural counterparts.

3.2.2 Structural Attributes

Structural attributes describe the physical structure of a property and the land parcel within which it is located. Intrinsically, these attributes represent the shelter afforded by housing and the physical investment by the owner. Thus it can be argued that they provide the greatest utility to the consumer, and hence have the greatest weight in a utility function (Bajic, 1984). Furthermore, compared to locational attributes, structural attributes are conceptually more tangible and can be more accurately perceived. For instance, the number of rooms in a property is far easier to measure than is street quality. Structural attributes can be categorised into two groups (Follain & Jimenez, 1985. pp. 95. Table 2). The first contains attributes that pertain to living space; the second to structural quality. Although attributes of living space have commanded much more attention in the literature, Kain & Quigley (1970a) have suggested the structural quality of the housing bundle has at least as much affect upon price, and that 'understanding the nature of the urban housing market requires a better grasp of these kinds of interrelationships' (pp. 545).

3.2.3 Locational Attributes

'[I]n the city, everything affects everything else' (Lowry, 1965. pp. 158 quoted in Harvey, 1973. pp. 58)

Locational attributes are measures of locational externalities. These have been defined as unpriced effects that affect the utility of others (Pinch, 1985. pp 8-9). They are unpriced in the sense that they are not paid for directly, but indirectly through housing purchase. They tend to be spatially concentrated in their impact upon the quality of people's lives and the value of their property. For this reason, externalities have been traditionally couched in terms of conflict between landuses that generate them and the residential location of different groups of people who may be affected by them (Cox, 1979). Pinch (1985) distinguishes between externality effects depending upon whether or not they impose a cost or benefit to the householder. These shall be discussed in turn.

3.2.3.1 Negative Externalities.

A negative externality is an effect that creates disutility to the consumer. For instance, a factory may pollute the air in a residential neighbourhood and cause disutility to the households of that neighbourhood through poor air quality, ill health, and cleaning costs for which the factory does not offer any compensation. In other words, the factory externalises the costs of production (Cox, 1979). Amenities which are associated with such costs are often termed noxious facilities and tend to have a negative affect upon house prices.

3.2.3.2 Positive Externalities

These are the opposite of negative externalities. They are benefits received as a by-product of an activity to whom the beneficiaries do not provide any direct payment. A good example of this is open space in an urban area which, although benefits all the residents of the urban area, has an additional benefit for the residents directly adjacent to it. As such, they tend to have a positive influence upon house prices. In some cases, a single amenity may emit both positive and negative externalities, such as a local shopping centre. This imposes negative externalities in the form of traffic congestion and noise pollution and positive externalities through convenience to the local residents.

3.2.3.3 Asymmetrical and Reciprocal Externalities

It is also possible to differentiate between asymmetrical and reciprocal externalities (Cox, 1979). With an asymmetrical externality, the producer and consumer of the effect can be distinguished. In the case of negative externalities, the producers imposes disutility and gains at the consumers' expense, such as air pollution from a factory. With positive externalities, the producer generates utility and the transfer of gain is in the opposite direction, as with public parks. In reciprocal relationships, the producers and consumers of externalities are the same. Traffic congestion, for example, is a negative externality produced by drivers and consumed by fellow drivers. They both produce costs for others and experience the costs imposed by others. In the case of positive reciprocal externalities, both producers and consumers benefit from each other. For instance, the residents in a neighbourhood may provide benefits for each other in the form of mutual assistance such as neighbourhood watch schemes. However, this asymmetrical and reciprocal dichotomy is rare, since most externalities are a mixture of both relationships.

3.2.3.4 The Effects of Externalities.

Most externalities are local in their impact, with a distance decay effect in their extent and intensity. Generally, households closest to the source of the externality will be affected the most, with the intensity of this effect diminishing with distance. A park will be of the most benefit to those households immediately adjacent to it, with this benefit diminishing rapidly with distance. Also, the larger the facility, the greater the intensity and range of these effects. Hence, large parks will have a greater and geographically wider influence on house prices than small parks. A major problem is how to measure the range of these effects. Many studies have used arbitrary thresholds around a facility to represent a catchment area, but this does not take into account the distance decay effects which will vary with the amenity (Pinch, 1985). In many cases the decay effect may not be a monotonic function of distance, since many externalities have both a positive and negative impact. For example, although a location next to a school may be beneficial with respect to accessibility, this benefit may be negated due to school related externalities such as noise and traffic. Therefore the distance decay function will be non-monotonic with the optimal location viewed as a trade off between the benefits of increased accessibility and the costs of proximity (Harvey, 1973).

Another problem is that the extent to which an externality is perceived as a cost or benefit will vary between individuals, and whether they use the facility or not. Air pollution from a factory will probably be perceived as less of a cost to those individuals who work at the factory than those who do not. This implies that externalities will be positive for some members of society and negative for others. In attempt to clarify this conflict, Dear (1976) defined 'user-associated' and 'neighbourhood-associated' externalities. Generally, 'user-associated' externalities benefits the consumer who may live beyond the neighbourhood within which it is located, but not the non-users who live in the neighbourhood itself, and vice versa. In addition, Knox (1995) distinguishes between 'public behaviour' neighbourhood externalities and 'status' neighbourhood externalities. The former concerns the effect of peoples behaviour in public, such as tidiness, quiet and sobriety, whilst the latter relates to the reflected glory of living in a distinctive neighbourhood. Hence, the affects of location on price will also depend upon the income of different groups, their demography and their social attitudes.

Section 3.3 Attributes Used in Previous Hedonic Models

Although much is known about the attributes that affect the price of housing, theory offers little guidance in determining which of these to include in the hedonic model (Ohsfeldt, 1988). This is demonstrated in the literature, where the lack of widespread agreement has resulted in a diverse range of variables entering the hedonic specification. Furthermore, far more attention has been paid to date to structural attributes than to locational ones (Cheshire & Sheppard, 1995). All previous models have contained similar structural attributes to the extent that they represent the majority of the variables in the model. Graves et al, (1988) have categorised the variables included in the hedonic model in terms of being either free, focused or doubtful. Free variables are those that are known to affect house prices, but are of no special interest in the study. These include certain structural and locational variables, such as floor area and distance to the city centre which may be of no particular interest to the researcher, but which will seriously bias the results if omitted. Focus variables are those variables of particular interest, and hence whose inclusion may vary from study to study. Doubtful variables may or may not affect the independent variable, but whose *a priori* omission may bias the results. Since the term doubtful variables have generally been applied to locational attributes, their importance in previous models has been very inconsistent, to such an extent that hedonic research can be classified on the basis of the inclusion of locational attributes. Hence, the following is an overview of previous research based on such a classification. This classification is also substantive in the sense that the treatment of locational attributes reflects the availability of spatial data.

3.3.1 Dwelling Specific Models

These models, also called location insensitive models (Ball, 1973), treat housing as essentially aspatial and disregard their geographic location as an unimportant factor in determining price. This suggests that households benefit from the structural attributes of the property only, and gain no utility from its location. Although this is theoretically and conceptually unsupported, structural attributes have been given prominence, particularly in early research (eg Cubin, 1970; Kain & Quigley, 1970a; 1970b; Wilkinson 1971; 1973b). The types of structural attributes used have varied, and to some extent seem to have been determined by data availability.

Following Follain & Jimenez (1985) categorisations, Table 3.1 is a summary of the most common structural attributes from previous research. Measures of living space have been reduced to lot size, floor area and the number of rooms, whilst structural quality is concerned with age, style and interior and exterior quality. Previous studies have argued that floor area and number of rooms give sufficient information about the size and structure of the dwelling (Lineman, 1980). Both attributes are relatively non-malleable and their supply is fixed in the short-run (Bajic, 1984). Also, to avoid overlap between lot size and size of the dwelling, outside lot area and internal floor area are often used instead of total lot size.

Table 3.1
A Summary of Commonly Used Structural Attributes

Classification	Attributes
Living Space	Interior Total Interior Living Space Total Floor Area Number of Stories Number of Rooms Number of Bedrooms Number of Recreation Rooms Number of Bathrooms Basement Attic Exterior Lot Size Off Road Parking Number of Garages
Structural Quality	Index of Dwelling Quality Age of House Presence of Full Insulation Brick Exterior Style of House Presence of Fireplace Double Glazing Air Conditioning Full Central Heating

Structural quality is often harder to measure and is more subjective and corresponds to the physical condition of both the interior and exterior. The structural quality attribute is often constructed from several questions about the state of repair of the structure. For instance, Ohsfeldt (1988) constructed an index based on a scale of 0-6, determined by questions on state of repair and basic facilities. However, since an index of six simply measures a house in an average state of repair with basic amenities, this is actually an index of disrepair and deficiency as opposed to one that measures a range of structural quality . To rectify this, studies often include measures of enhanced quality such as central heating, double glazing and insulation to capture variations in structural quality of dwellings which do not have basic structural problems.

Wilkinson (1973a) and Ball (1973) may have been mistaken to describe these models as purely dwelling specific and locationally insensitive. Theoretically, a property's structural attributes and its location within the city are related, since they reflect the growth of the urban structure (Muth, 1969). This implies that an element of location will be inherent within the physical structure of the property. For example, although the age of the property, the size of building parcels and the patterns of tenure are all structural attributes, they will tend to vary systematically across urban space, reflecting historical growth patterns (Batty & Longley, 1987). Thus, such attributes may also proxy locational measures. This is indicated in studies such as Cubbin (1970) and Kain & Quigley (1970a), which revealed a high degree of multicollinearity between structural attributes, and the results suffered from spatial autocorrelation.

3.3.2 Location Specific Models

In most studies, locational externalities have been conceived in terms of relative and fixed locational attributes. Relative locational attributes are measures that reflect the externalities of the local neighbourhood and are unique to an individual property, such as street quality. Fixed locational attributes capture the location of a property with respect to the whole urban area, and pertains to some form of accessibility measure, typically access to the CBD. Taken together, relative and fixed locational attributes are said to equal the total utility of location (Krumm, 1980). Krumm argues that the important distinction between fixed and relative locational attributes is that a household may only change the level of the former through migration, but may affect the latter through the use of resources at that location. Hence relative locational attributes are produced, to some extent, by all households in the neighbourhood. Although this dichotomy has caused some disagreement with respect to which attributes fall into which category, it has generally been accepted and hence, Follain & Jemeñez (1985) have classified locational attributes from previous studies into categories of (fixed) accessibility and (relative) neighbourhood quality measures.

3.3.2.1 Accessibility and Fixed Locational Attributes

Accessibility has been the most fiercely contested and the single most important measure of location in hedonic house price models. Traditionally, it has represented the measurement of the bid-rent curve as proposed in the micro-economics literature. It was this literature that was the stimulus behind much of the initial hedonic work, with Muth's (1969) pioneering work on estimating the rent gradient for Chicago being of particular significance (Dubin & Sung, 1987). Therefore it is not surprising that most of the hedonic research has been built upon the monocentric models of Alonso (1964), and later Evans (1973). These espouse the importance of the city centre as the major factor influencing land values, with the resulting bid-rent curve translated into a negative house price gradient. However, more recent work in this area has acknowledged the complexity of accessibility and residential location, and has suggested that the monocentric model be substituted for a multi-centric or polycentric model (Gordon et al, 1986). Such a model describes a city as having more than one, and usually several, population and employment centres, although the CBD is still envisaged to be the most prominent one (Griffith, 1981). Sanchez (1993) quotes changes in economic conditions, transportation costs, technology, and social patterns as giving rise to the sort of urban morphologies and land value distributions that are increasingly less 'center orientated' (pp. 455). It has often been claimed that these models are more representative of modern (US) cities, resulting in the hypothesis of an urban area with multiple house price gradients (eg. Heikkia et al, 1989; Waddell et al, 1993). However, since much of the theoretical work is vague regarding the definition of non-CBD centres, the exact nature of what is being measured is questionable. For instance, although a polycentric model may depict an urban area as having several centres, each centre may have a different function, and as such, a varying degree of influence on the urban area (Griffith, 1981). This is in contrast with the monocentric city which has a single, well defined multi-functional focus that is hypothesized to have an affect across the entire city. This vagueness is also exacerbated by a lack of complementary empirical work, so that accessibility measures within a polycentric context have little to go on. These issues will now be expanded upon in the next section.

3.3.2.2 Measurement of Accessibility

I. The Monocentric Urban Model

The majority of hedonic house price research has been theoretically underpinned by the monocentric model, which proposes the existence of a negative house price gradient from the city centre reflecting the trade off between house price and declining accessibility. Cheshire & Sheppard (1995) argue that if locational attributes are appropriately measured, then monocentric models can perform well in a UK context. However, as will be discussed, this may not necessarily be true for modern US cities.

Various different measurements have been implemented to capture the affect of accessibility with a equally differing results. Straight-line distance to the CBD has been the usual measure, with a linear or semi-log functional form, with a recent shift to route distance, although there is no strong argument that this is an improvement (Cooley et al, 1995). Physical distance is not the only measure of accessibility that has been used. Accessibility in terms of journey to work time, travel time and monetary loss have also been considered (eg. Wabe, 1971; Bajic, 1984; Sanchez, 1993). In terms of the complex and numerous permutations reflected in the modal split found in most urban areas, measuring accessibility in terms of merely physical distance or one form of transport will undoubtedly oversimplify the problem. Also travel time may vary with hour of day and day of the week.

The functional form of the accessibility measure need not be smooth, or continuous since structural features such as arterial roads, rivers and railway lines can all distort accessibility. Moreover, transport systems, and so transport costs, are not necessarily the same across an urban area, so the accessibility function may vary with direction from the city centre (Cheshire & Sheppard, 1995). Such accessibility measures have been explored by Dubin & Sung (1987) and Waddell et al, (1993). They both divide the urban area into sectors and use dummy variables to allow shifts in the slope of the price gradient which would allow a complex functional form to be represented in a simple manner, and thus avoiding specifying the functional from in an *a priori* manner. Of course, the estimation of the functional form will depend in part on the dummy variables used. Waddell et al (1993) used arbitrary distance intervals, whilst Dubin & Sung (1987) hypothesized the location of breaks in the price gradient. Both conclude that these accessibility measures performed better than when standard functional forms, such as log of distance, were used.

Distance to transport routes, such as main roads, railways and bus routes have been an important feature of accessibility measures, since it is hypothesized that proximity will increase property values because of increases in accessibility and decreased transportation costs (Forrest et al, 1996). Of course, with respect to motorways and railways, it is distance to intersections and stations respectively that are of interest. On the other hand, transport systems generate negative externalities such as noise and air pollution and congestion which may eliminate any locational advantages of proximity (Sanchez, 1993). For instance, Waddell et al (1993) demonstrated that the highway proximity gradient in their study was non-linear, due to the associated negative externalities depressing property values close to a highway, but with the values increasing a short distance away once these externalities effects had become minimal.

However, previous empirical studies provide no consensus on the magnitude, or in some cases, even the existence of a price gradient. It is quite common for empirical studies to fail to find a statistically significant between city centre access and price. Several studies have even reported statistically significant accessibility influences of the wrong sign, that is, price increasing with distance. These conflicting results are often attributed to the inadequacy of traditional accessibility measures (eg. Goodman, 1979; Heikilla et al, 1989). In an overview of early work, Ball (1973) particularly emphasized this lack of consensus, and himself attributed it to poor data and the variety of measures used. However, he also suggested that there was no reason to assume that all cities should produce the same results, since the price gradient should be viewed within its historical and geographical context. For instance, Ball & Kirwin (1977) suggested the historical geography of the housing stock in Bristol, with affluent suburbs close to the city centre, resulted in the overall importance of accessibility in determining prices being smaller than anticipated. These contradictory results have since been blamed upon the theory of urban residential location that underpins it, and in particular the trade-off mode: 'It is not clear that the trade-off between commuting and land rent plays any significant role at all in location decisions' (Hamilton, 1982; pp. 1050).

II. The Polycentric Urban Model.

Cities rarely have a simple monocentric structure, since employment and amenity centres are often located outside of the city centre, and this may cause the house price gradient to be complex, and undermine the significance of the city centre price gradient. Indeed, commentators such as Jackson (1979), Hamilton (1982) and Dubin & Sung (1987) have commented upon the implicit and often unjustified assumption of the monocentric city, and have criticised the negligible attention paid to the possibility of other influential sub-centres. Ball & Kirwan (1977) view the monocentric city as a 'major shortcoming' in a significant amount of hedonic research, whilst Hamilton (1982) asserts that the 'widespread acceptance of the class of monocentric models seem to rest more on its intrinsic plausibility than on any demonstrated ability to withstand empirical scrutiny' (pp. 1035). Furthermore, Bender & Hwang (1985) point out that the monocentric city is only a special case of the standard urban model. They argue that although the standard urban model proposes a negative house price gradient from the city centre, it also allows for the existence of secondary employment centres outside the CBD, which 'is not only intuitive but also empirically relevant' (pp. 91). Failure to control for commuting time to these secondary employment centres will tend to positively bias the coefficient of accessibility to the CBD. This, they suggest, probably explains the 'luke warm support' for a negative house price gradient in previous research (ibid). Hence, the urban area may be visualised as a price surface with a global maximum at the CBD and local maxima at the secondary employment centres (Jackson, 1979). This multiple-access hypothesis rests on the idea that the value of a property is determined in part by distance to each of these centres.

Empirical research on the nature of property prices within a polycentric urban context has so far been comparatively scarce, with the few exceptions including Jackson (1979), Bender & Hwang (1985), Dubin & Sung (1987), Heikkula et al (1989), and Waddell et al (1993). However, the definition of a secondary centre in such research has been vague. Typically, they refer to concentrations of non-CBD employment, but other urban amenity centres have also been considered. These have included shopping centres, hospitals, airports, and cultural centres such as universities. In fact, these features were hypothesized by Muth (1969) as the cause of local peaks in his estimated rent gradient. But it is difficult to determine how the traditional concept of accessibility, the measurement of the bid-rent curve, can be applied to some of these proposed secondary centres. The assumption is that, similar to the CBD, the secondary centres provide a large bundle of public services which are capitalised into property values. It can be hypothesized that this may be the case for employment centres, since these will usually generate a larger demand for labour in the local vicinity, relative to the CBD. Sanchez (1993) has further argued that, if the secondary employment centre is large enough, house price decrease will be affected by it, regardless of proximity to the CBD. However, it is questionable whether the other proposed secondary centres, such as airports and hospitals, will have the same effect. For instance, Waddell et al (1993) showed that the price effect of major secondary employment centres remained significant over a much larger area than more localised amenities such as airports and retail centres. Besides, as will be explained in the next section, some of these features have also be defined in terms of neighbourhood externality effects. Hence, there seems to be a confusion in the literature between accessibility effects and the proximity effects of externalities.

3.3.2.3 Multiple Accessibility Measures

There have been two main methods of estimating accessibility within a polycentric urban model. The first has measured the effects of multiple centres by estimating an accessibility trend surface (Jackson, 1979). Such a model holds housing attributes constant across an urban area, but allows the price of land to vary spatially as a result of demand for more accessible sites. The power of the polynomial of the trend surface represents the accessibility surface. For instance, a quadratic approximation represents a surface with a single maximum value for accessibility; the monocentric city. More complex surfaces with multiple local peaks can be represented by increasing the degree of the polynomial. Jackson advocated using a combination of R-square and F-tests to ascertain which degree of polynomial estimates the accessibility surface the best. The resulting surface can then be mapped. The advantage of such a procedure is that the secondary centres that have a significant influence upon house prices do not have to be specified *a priori*. In the case of Milwaukee, Jackson discovered that the price surface was centrally located with respect to manufacturing employment, and then decreased in all directions, but increased again towards secondary centres such as the university and peripheral manufacturing areas. The contour lines followed fairly regular concentric rings, with distortions caused by the transport system.

The second method is to identify secondary employment centres *a priori*, and estimate price gradients using traditional accessibility measures. The usually procedure was to experiment with various hypothesized secondary centres and select those that produced the optimal results based upon R-square values and statistical tests. However, Heikkilla (1988) argues that it will be intuitively expected for two or more distance measures to be collinear when confined to a plane, and hence multiple accessibility measures may be subject to problems such as multicollinearity. Also, there is no reason to expect every secondary centre to have an effect upon every house in an urban area (Jackson, 1979), and so the urban area was partitioned and each property allocated to its closet centre. The results of the multiple-access studies share the same diverse results as the earlier monocentric work. With the exception of Waddell et al, (1993) study of Dallas, a common result was the surprising failure of the CBD to exert a dominant influence on the overall house price gradient. In the case of Baltimore, Dubin & Sung (1987) concluded that 'the CBD appears to behave like the other [secondary] centres: it has an impact, but this effect is limited to a relatively small area' (pp. 204.) Similar results were concluded for Milwaukee (Jackson, 1979) and Chicago (Bender & Hwang, 1985) . In the case of the latter, it was concluded that 'in areas relatively close to a particular employment centre, accessibility to that centre is the dominant accessibility variable influencing price' (pp. 102), and that Chicago could be best characterised as consisting of a major monocentric city and two minor monocentric 'cities' with overlapping boundaries. Furthermore, Heikkila et al, (1989), in a study of Los Angeles, found that the CBD price gradient became positive and statistically insignificant once distance to multiple employment centres were explicitly included in the model. They concluded that the total lack of influence of the CBD was due to Los Angeles being a special case in terms of dispersed employment and population. This may be the case, but Waddell et al. (1993) pointed out that the study estimated a highly significant and positive coefficient for age of dwelling which was troubling since this is counter-intuitive. They suggest that this is indicative of collinearity between age and distance from the CBD, as discussed with dwelling specific models, and advocated the use of dummy variables.

However, the conclusions from these studies are very similar and suggest that maybe it is misleading to assume that accessibility to the CBD is of sole importance in all urban areas, particularly those of a large, dispersed character, and that it could be that the price of housing could be influenced by attributes of a more localised nature.

3.3.2.4 Neighbourhood Quality and Relative Locational Attributes

"Obviously the value of land in any city is not a function of distance from the city centre alone: there are other exogenous variables" (Evans, 1973; pp. 60)

Most studies have acknowledged that the neighbourhood within which the dwelling is located is an influential factor affecting house price. Indeed, Muth (1969) discusses several factors that affect house prices, other than structural attributes and distance from the city centre, and these in general can be described as measures of neighbourhood quality. Studies have tended to account for neighbourhood quality by either explicitly including neighbourhood attributes or by stratifying the sample using the neighbourhood as the basis. But, as with accessibility, there appears to be little agreement upon how neighbourhood quality should be measured and its inclusion seems to be based upon data availability. Hence, Ball (1974) comments that neighbourhood quality is used to cover a large, ill defined set of influences on house prices. Furthermore, it has been argued that neighbourhood quality as it stands is an unobservable attribute that can only be measured indirectly by the use of proxy measures (eg. Davies, 1974). However this is debatable since many aspects of neighbourhood quality are tangible and have been quantified, such as in The English Housing Condition Survey (e.g. 1991). But it is correct to argue that much hedonic work have relied upon proxy measures instead of direct measures of neighbourhood quality. Graves et al (1988) described these measures as doubtful variables since being a proxy to a true measure, it is not known whether they affect house price.

Despite this, Dubin & Sung, (1990: pp. 98) have classified measures of neighbourhood quality used in previous research by three broad categories; measures of local public amenities, measures of the socio-economic status of the neighbourhood, and measures of neighbourhood racial composition. This classification illustrates well the broad measures used to capture neighbourhood quality, with the use of proxy measures emphasized by the lack of a category for explicit measures of environmental quality. In comparison, Mingche & Brown (1980) argue that many studies have included few, if any, 'location-specific' attributes and advocates measures of the 'micro-neighbourhood' that are defined in terms of aesthetic attributes, pollution levels and 'proximity', by which they mean accessibility to local amenities. These attributes can be regarded as direct measures of neighbourhood quality, and are more analogous to the definition of locational externalities. However, as will be argued in the next section, the social and racial composition of a neighbourhood may still be influential, in addition to the quality of the environment. As Wilkinson (1973a) summarised: 'Neighbourhoods can be measured in terms of the characteristics of the dwelling, the people, and the physical and social amenities which comprise them' (pp. 76)

3.3.2.5 Measures of Neighbourhood Quality

I. Public Amenities

Attributes relating to public amenities are generally the most straightforward to measure and interpret since they are principally regarded as direct measures of neighbourhood quality and are easily quantified. Generally the better the quality of the service, the more highly valued it is and so is positively capitalised into house price. However, Dubin & Sung (1990) concluded that they discovered that services were relatively unimportant in contrast to socio-economic and racial composition of the neighbourhood. They admit that this finding was surprising in light of the emphasis on public provision in the literature. Most studies are invariably concerned with the quality of local schools, and hence common factors include the pupil/teacher ratio and average examination results (Fortney, 1996; Cheshire & Sheppard, 1995; Herrin & Kern, 1992). Other public amenities have included the amenities provided by public parks, (eg. McLeod, 1984;), golf courses (Do & Grunditski, 1995), and the availability of local shops (Powe et al, 1995; Lineman, 1980). Also included in the public amenity category are measures of local property tax rates and local government jurisdictions, both of which can have a bearing upon the quality and cost of public service provision.

II. Socio-Economic Status

Measures of the socio-economic status of a neighbourhood are less tangible, and have been classified as 'doubtful variables' (Powe et al, 1995). Typically measures have been constructed from census variables that relate to income levels, education, age and car ownership. Other indices include ACORN classifications in a UK context (eg Forrest et al, 1996; Collins & Evans, 1994) which are marketed as 'pen portraits' of the attributes of an area based upon the characteristics of the residents in them (Longley & Clarke, 1995). The crux of the uncertainty is whether the presence of high income households increase the value of certain neighbourhoods or whether certain neighbourhood features attract high income households (Sanchez, 1993). If the former is the case, then socio-economic status can be thought of in terms of a direct measure of neighbourhood quality. Members of high socio-economic groups are thought to be more desirable neighbours since they value the quality of the local environment greater than those in lower social groups and as a result may be prepared to make larger investments to maintain that quality. This is supported by

Knox (1995), who suggests that environmental quality is closely tied to patterns and processes of investment and disinvestment and of social segregation. Alternatively, in the case of the latter, neighbourhood quality is regarded as being income elastic, so that it is likely to be given greater weight by higher income groups who are attracted by such attributes. In this case, socio-economic status represents a proxy for other attributes of neighbourhood quality such as low levels of air pollution, a low crime rate and high aesthetic surroundings. However, it is unlikely that socio-economic status will fall precisely into either category. It can be argued that high income households can afford to live in attractive residential areas, and if these areas are in short supply, will out-bid households in lower income groups who must settle for cheaper housing, often in less attractive surroundings. But at the same time, the clustering of income groups into specific areas will tend to create many of the neighbourhood features that each group finds favourable, and this will attract the same kinds of households. Therefore, it can be argued that socio-economic status is in fact a measure of 'public behaviour' and 'status' neighbourhood externalities, as described by Knox (1995), and should be included as a direct measure in addition to, and not to the exclusion of, other attributes of neighbourhood quality.

III.Racial Composition

The case for racial composition is even less clear. Again, there is a disagreement over whether race is a direct measure or proxy for neighbourhood quality. The case for a direct measure argues that discrimination against racial minorities reduces their access to housing and consequently causes the price of housing to be higher than in white neighbourhoods (Berry, 1976; Schnare, 1976). A similar argument follows, namely that people of the same race prefer to live in the same neighbourhood and hence houses in segregated neighbourhoods are more in demand than houses in integrated ones, and this will be capitalised into its price (Daniels, 1975). This has been called the 'taste for segregation' model (Bender & Hwang, 1985) but has encountered difficulty with respect to the definition of segregated neighbourhoods. If, on the other hand, racial minorities prefer white neighbours then the price of a house would increase as a monotone function of the percentage white (Waddell et al, 1993). However, if the racial composition of a neighbourhood simply reflects other characteristics such as socio-economic class, income and depressed surroundings then race is merely a proxy for neighbourhood quality. However, Dubin & Sung (1990) discovered that both race and socio-economic class were important, and that neither alone could sufficiently explain house price variation. The majority of studies into the influence of racial segregation on house price have been in a US context. It is arguable that in a UK context, racial composition may not be as significant a factor in the majority of urban areas due to greater social, cultural and economic heterogeneity within the majority of cities.

3.3.2.6 Environmental Quality

According to Richardson (1976), the quality of the environment is one of the most important determinants of a households location, and thus the price they are prepared to pay for a property. Under the Mingche & Brown (1980) classification, environmental quality can be regarded in terms of the aesthetics of the local area, the pollution levels and also proximity to local amenities.

I. Aesthetic Measures

Environmental quality is often associated with open space, such as fields, parks and beaches. Although such features may have an amenity value, such as for leisure, a view of the feature may also be perceived as a benefit and hence will be capitalised into the price of a property. As Gillard (1981) argues: "Even when a park may not be used for recreation because of crime problems, it may still be valued for aesthetic reasons by residents with a view of the park." (pp. 217)

This can be thought of as the aspect of a property. Previous research into aspect has been particularly interested in features such as river views (Lansford & Jones, 1995; McLeod, 1984, Darling, 1973), forestry (Tyrvainen, 1997; Garrod & Willis, 1992) and shore line (Brown & Pollakowski, 1977). McLeod (1984) discovered that river views were particularly important, and had a greater influence than a view of a park. These results were supported by Lansford & Jones (1995), Darling (1973) and Gillard (1981), although were disputed by Davies (1974) and Brown & Pollakowski (1977). Of course, it is not expected that results from different cities should be identical. As Gillard (1981) argued, the price is not dependent upon the intrinsic worth, but the supply relative to demand. In some urban areas, aesthetic views may be so abundant that they may be regarded as a free good. Other measures of aesthetic quality common in studies have included topography, with elevated areas being more desirable, and explicit measures of street quality such as condition of roads and pavements. A further measure includes population density, although this may be a proxy for attributes such as open space. Alternatively, certain views can have a negative effect upon property values. In particular, industrial, business and transportation land uses can have a negative effect upon property prices with respect to aesthetic qualities (Powe et al, 1995).

II. Pollution Measures

Pollution, and specifically air pollution, has been the focus of a great deal of hedonic research. Perhaps the most influential study is Harrison & Rubinfeld's (1978) study of air pollution in Boston caused by traffic and industry. Similar studies into air pollution have included Ridker & Henning, (1967), Lineman, (1980) and Palmquist, (1984). The other important pollution measure has been noise pollution, particularly with respect to airports (eg. Levesque, 1994; Collins & Evans, 1994; McMillian et al., 1980; Goodman, 1979). In a summary of the evidence concerning the price effect of airport noise from a variety of studies, Nelson (1980) concluded that excessive noise depresses property values. Moreover, this effect appears to exhibit a considerable similarity in a variety of cities at different times. Other source of noise pollution, such as that caused by roads and railways, have also been investigated (eg Cheshire & Sheppard, 1995; Hughes & Sirmans, 1992; Krumm, 1980).

III.Proximity Measures

These correspond to measures of locational externalities which Mingche & Brown (1980) have argued have been lacking in most studies. The attributes in which proximity is regarded as being significant are access to public amenities such as schools and shops, non-residential activities such as industrial sites and open space. Since proximity can have both a positive and negative effect, they advocate the use of a non-monotonic distance function. This was corroborated by Waddell & Berry (1993), who concluded that a non-linear function was discovered for the amenities they measured, and can be expected where access is valued, but where immediate negative externalities over-ride the gain increased from proximity. The amenities included major and minor shopping centres, universities, hospitals and airports, and in accordance with externality theory, different shaped distance decay functions were estimated for each amenity, to reflect the fact that the effects of different amenities vary in their range and magnitude on property prices. For instance, Waddell & Berry (1993) discovered that in Dallas, major retail centres had only a slightly negative slope, and hence positive effect, in the first half mile, reflecting the capitalisation of access in price, but also the negative effects of immediate proximity. Once outside the influence of the negative externalities of the shopping centre, the slope increased in magnitude between two to five miles before decaying. Minor shopping centres had a similar distance decay function, but were smaller in magnitude and decayed more rapidly.

However, proximity has traditionally caused controversy within hedonic studies, due to conflicting results in both general and specific research into its effects. Hence, in accordance with theory, Do et al, (1994) found that neighbourhood churches had a significant negative effect upon house prices, with this impact decreasing with distance from the church. They attributed this to negative externalities associated with congestion and noise pollution. More generally, Kain & Quigley (1970a) found by factor analysis that the presence of commercial and industrial uses on a street had the expected negative effect upon property values. Similarly, Stull (1975), in a study of Boston, found that the median value of single-unit owner-occupied homes were negatively related to both the proportion of land devoted to multiple-family use and that devoted to industrial land use. Also, Stull found that, as the proportion of commercial land in a neighbourhood exceeds five percent, the values of single-family homes tend to fall. However, Grether & Miesszkowski (1980) discovered that the results of proximity in New Haven were mixed. Industrial and public housing areas had a significant, negative effect whilst minor commercial centres have relatively little effect. They concluded that non-residential land use *per se* has no systematic effect on housing values, and that landuse externalise may be very localised in their effect, so that they are a 'next-door' phenomena. A similar conclusion was reached by Li & Brown (1980), after distinguishing between the value of access to non-residential land uses from the externalities generated by their uses.

3.3.3 Conclusion

To conclude this section, it should be clear that locational attributes have been problematic in previous research. Firstly, the distinction between fixed and relative locational attributes is not clear, with amenities such as hospitals and airports being categorised in terms of both accessibility and local proximity measures. Secondly, it should be evident that neighbourhood quality is a tangible and quantifiable feature, although as previously discussed, in the majority of studies it has been inferred by measures of socio-economic status and racial composition

Section 3.4 The Nature of Data in Hedonic Models

3.4.1 Problems with Data Measurement

The above discussion has highlighted several themes related to the data used in previous research. Firstly, it should be evident that there are marked differences between the scope and quality of measures of structural attributes when compared to locational ones. Whilst the potential set of structural attributes is limited and well researched, 'the set of available neighbourhood [attributes] is nearly infinite' (Cheshire & Sheppard, 1995; pp. 24). They argue that this is both because locational data are more difficult to collect and because it is less obvious, *a priori*, which locational attributes are relevant in determining prices. This has resulted in two problems. Firstly, Butler (1982) infers from this that any estimate of the hedonic price function will be misspecified to some extent because some of the relevant variables will inevitably be omitted, whilst Lineman (1980) comments that this missing variable bias is a particular problem for locational attributes. An omitted variable will violate the error term since it will capture the effects of the missing variable. In addition, this may also bias the included variables, since they may compensate for the omitted variable which can lead to erroneous interpretations of the parameters. For instance, although Sanchez (1993) used miles of street per square mile of land in each census tract as a measure of accessibility, he points out that this could also be a gross indicator of housing unit density and lot sizes in each tract.

Secondly, the problems of data availability have led to the use of surrogate or proxy measures. This has been illustrated with the case of socio-economic class, which has frequently been used as the only measure of neighbourhood quality. Also, such proxy measures are often constructed from aggregated data. In the majority of cases, the quality of the data collected for these aggregated spatial units are 'notoriously poor' (Anselin, 1988b; pp. 283), typically census data aggregated to wards or enumeration districts. The problems caused by poor neighbourhood quality attributes on measures of accessibility have been well documented, such as the research by Herrin & Kern (1992) which showed that better neighbourhood measures yielded significant improvements in the accessibility parameter. More importantly, the lack of good data measurement can cause standard econometrics to fail in numerous ways when applied in a spatial context. As discussed in *Chapter Two*, the presence of spatial aggregation, spatial externalities and spillover effects will separately, or in combination, affect the properties of the hedonic price function and statistical tests.

A general rule of thumb in the basic hedonic literature is that the more variables that increase the adjusted R-squared, the better. Thus, it is common to find more than twenty attributes in the hedonic price function although this can lead to severe problems with multicollinearity, especially if poorly specified data are used. Multicollinearity can be expected due to the inter-relationships between structural attributes and locational attributes (Ozanne & Malpezzi, 1985). Such was concluded by Powe et al (1995), who claimed that the high correlations between amenity variables, socio-economic variables and structural variables was the single greatest problems in their empirical research. Structural attributes will tend to be collinear because of the relationships between house size and house structure. Large houses will tend to have more bedrooms and recreation rooms for instance, than smaller houses. Inter-relationships between variables are more problematic for locational attributes. The relationships between socio-economic class and neighbourhood quality have already been discussed. Similarly, accessibility measures have been plagued by multicollinearity with neighbourhood quality variables. This factor has been blamed for poor and counter-intuitive results, which has resulted in the cries of 'what happened to the CBD-distance gradient?' (Heikkila et al, 1989). For instance, Goodman (1979) argues that negative externalities emanating from the CBD, such as noise and air pollution, will be negatively correlated with distance and this may cancel out the affect of accessibility since if 'the relationships are of similar magnitudes with opposite signs, it is possible for the distance term to be insignificantly different from zero and considered unimportant' (pp. 327). Moreover, multicollinearity between structural and locational attributes have long been a cause for concern. Kain & Quigley (1970a) appreciated that the quality of the neighbourhood is to some extent influenced by the dwelling stock, and commented that the difficulty in separating the two was 'perhaps the most vexing problem encountered in evaluating the several attribute bundles of residential services' (pp. 533). More recent examples includes the previously discussed critique of Heikkila et al (1989) work on Los Angeles by Waddell et al, (1993), who concluded that collinearity between age and distance variables had resulted in the counter-intuitive result for accessibility.

However, despite the fact that multicollinearity is a common problem of hedonic models, it is 'one which is often conveniently ignored' (Garrod & Willis, 1992a; pp. 65). With the few recent exceptions, such as Powe et al (1995) and Forrest et al (1996), explicit tests for multicollinearity have been lacking. Powe et al (1995) advocated the use of the variance inflation factor (Pindyck & Rubinfeld, 1991), and omitted the variables that were shown to have high partial correlation coefficients. Forrest et al, (1996) used the Klein test that had been commended by Maddala (1992) as being the most rigorous means of assessing multicollinearity. It is more usual for either multicollinearity to be dismissed as unimportant, or the model is tested for robustness by omitting variables suspected of causing multicollinearity and the model re-estimated.

3.4.2 The Spatial Resolutions of Data in Hedonic Models

The distinction that most research has made between fixed and relative locational attributes may not be helpful. The concept of locational externalities blurs this distinction, and such was Minchge & Brown's (1980) argument when they advocated specific measures of the micro-neighbourhood. By doing this they supported a more precise specification of locational externalities; that is, one which captured the magnitude and distance decay nature of the externalities. This cannot be achieved with the use of poorly specified 'blanket measures', which are typical of the types of proxy data that have been used. However, it is not the case that improving solely the quality of the data will improve accuracy of the models. A problem also lies in the fact that the attributes are strongly related to each other, and improved data may not resolve this problem. A possible method to overcome this problem is to use the expansion and multi-level specifications that were described in *Chapter Two.* These specifications acknowledge that some of the attributes are inextricably bound together, and takes this into account when the data are modelled. Moreover, the concept of multi-level location infers that locational externalities will operate across different spatial scales. For instance, accessibility externalities affect a wider area than neighbourhood or street externalities. This is an important concept, but one which appears to have been neglected in the literature. Instead, as was argued in *Chapter Two*, property prices have tended to be viewed as varying continuously in one dimension across urban space rather than over different spatial scales. Hence, it may be possible to overcome the data problems if the structure of property prices were to reflect the multi-level nature of urban space, and externalities were allowed to operate at different spatial resolutions. This was illustrated in *Chapter Two* at two levels: the property level and the submarket level. However, a street level may be inserted between these two, such that a hierarchy of three resolutions may be conceptualised, that intrinsically captures the externalities that operate within a local housing market. These are summarised in Table 3.2. The externalities in each level can be differentiated by how they are affected by the activities of households and the attributes of property, and the range and extent of their influence. Hence, since fixed locational attributes, such as accessibility to the CBD, and relative locational attributes, such as proximity to non-residential landuse, are unique for each property, they can be conceived as operating at the property level. However, locational attributes, such as street quality are

Table 3.2.
A Multi-Level Conceptualisation of Locational Externalities

Property Level Externalities	Accessibility to CBD Accessibility to Major Non-CBD Centres Motorway Exits Railway Stations Shopping Centres Suburban Employment Centres Proximity Measures to Non-residential Landuses Parks Schools Industry Commercial Local Shops Recreational Centres Cultural / Educational Centres
Street Level Externalities	Street Environment Class of Street Street Quality Non-residential Activity School Catchment Areas
Neighbourhood Level Externalities	Housing Density Proportion of Non-residential Landuse Proportion of Open Space Quality of Local Amenities Social Composition Racial Composition Prestige / Desirability

influenced by the activities of residents in the street and hence can be regarded as a street level externality. This may also be the case for local amenities, such as parks, if the effect of their proximity is localised. Relative locational attributes at the neighbourhood level are those externalities are that effect prices across wider areas. For instance, the effects of racial and social composition of a neighbourhood may be seen as operating at this level, as well as more amorphous concepts such as desirability. An externality may also operate at more than one spatial scale. For instance, the externality effects of local amenities such as a school, may operate at the property level with respect to issues of proximity, and also at the street level with respect to the catchment area. Both these two effects will influence property prices. This is discussed in more detail in *Chapter Five*, whilst the concept of how structural attributes and locational externalities should be modelled is considered in more detail in the next section, which examines how Geographic Information Systems (GIS) can aid the estimation of hedonic house price functions.

Section 3.5 GIS and the Analysis of Hedonic Models.

"The true potential of Geographical Information Systems lies in their ability to analyse spatial data using the techniques of spatial analysis." (Goodchild, 1988; pp. 76)

3.5.1 Introduction

It should be clear from the discussions in this and the previous chapter that one of the main problems encountered with hedonic house price models is the treatment of locational data, whether in terms of modelling of geographic space or in the measurement of locational attributes. These issues are representative of spatial data analysis in general (Anselin and Griffith, 1988) and can be linked to both the disregard of the potential problems posed by spatial data, and of the availability of well defined spatial data which has historically been poor. However, during the past decade, advances have been made which have gone some way to rectify these problems. Firstly, as was discussed in *Chapter Two*, the spatial nature of the hedonic house price function has been acknowledged in the new expansion and multi-level specifications. Secondly, there has been a marked improvement in the availability of spatial data at a much finer resolution than has previously been available. Thirdly, there has been the continued improvement of GIS, which can now store and manipulate much larger volumes of data at greater speeds and efficiency. The first issue of model specification has been dealt with in the previous chapter. The next chapter will discuss the types of data available for such analysis. Hence, the remainder of this chapter will discuss how GIS can improve the estimation of hedonic data.

3.5.2 GIS and Hedonic House Price Research

It would appear that GIS is an ideal medium to approach hedonic house price research for several reasons. It is capable of organising and managing large spatial datasets, such as those used in hedonic house price studies. Moreover, a GIS can handle these data at various spatial resolutions, such as at the level of the individual property and neighbourhood, which is important in the context of this research. A GIS also provides a valuable platform for spatial analysis, particularly with respect to the distance and proximity measures that have caused controversy in previous work. Finally, a GIS can aid the visualisation of the spatial data and map the results of the modelling. However, it would be easy to overstate the effectiveness of GIS in aiding such analysis, as has been done in the past, and hence some of these issues are considered in more detail below.

3.5.3 GIS and the Data Environment

Housing attributes represent a host of physical and socio-economic variables that are available at different resolutions. Hence, whilst structural attribute data are generally available for the individual property, it is more typical for locational attribute data to be aggregated at a higher level, such as census areas. A major defining attribute of a GIS is its ability to make such diverse data sets compatible (Flowerdew, 1991). A GIS can achieve this because it treats the attribute data of an object, and its location in geographic space, as two separate entities. By storing these two types of information separately in a database system, and allowing interaction between them, the data can be manipulated on the basis of either geographic location or attribute value. This flexibility is the power that underpins GIS. Furthermore, since a GIS is capable of storing these data at different scales, data can be geo-referenced and aggregated at various spatial levels. This is important if externalities are perceived as operating at different spatial scales, such as at the street level or the neighbourhood level. Thus, it would appear that a GIS is an ideal environment for storing and integrating the different types of data used in hedonic house price studies. But this situation is not as simple as this, since it is important to appreciate the nature of data within the GIS, particularly with respect that areal data.

Martin (1996) argues that because GIS is at least one step removed from reality, it is necessary to carefully consider the nature of geographic objects, as these are crucial to any subsequent use of GIS in answering geographical questions. He suggests that a model of the data environment should form the context within which any attempt to build a conceptual model of a GIS should sit. However, existing theoretical models of GIS have tended only to address the functions and component parts of the systems, without reference to the underlying data model, and its relationship to the real world. A fundamental distinction in areal data is that between data collected for artificial and natural areal units. Typically, physical data, such as area of parkland or length of street, are collected for well defined natural units. These units are natural since their boundaries are a meaningful spatial representation of the data in question. For instance, a street is a meaningful spatial object for data collected at street level. In contrast, socio-economic data tends to be collected for artificial areal units, that may have very little relation to the underlying population. It has been widely acknowledged, for example, that census geographies often bear little resemblance to the underlying population they are attempting to represent. Hence, an important consideration of the data environment is that any spatial study heavily depends upon the 'nature and intrinsic meaningfulness' of the spatial objects studied, particularly with respect to socio-economic data (Martin, 1996. pp. 54). The scale of the analysis is very important in GIS, and holding an appropriate model of the geographic world is fundamental to any form of GIS-based analysis. Moreover, the variety of basic units has presented a major obstacle, especially with respect to data collect for artificial areal units, as the patterns apparent in the data may be as much due to the nature of the collection units as to the underlying phenomena, and there is no direct way of comparing data collected for differing sets of areal units (Flowerdew & Openshaw, 1987). This is known more generally as the modifiable areal unit problem

3.5.3.1 The Modifiable Areal Unit Problem

This is a key problem in the manipulation of spatial data, and it is actually two distinct but closely related problems (Openshaw, 1984). There is the scale problem which is related to the level of aggregation of the data, and there is the aggregation (or zoning) problem, which is related to the fact that much of the data used in spatial analysis are based on areal units that do not have intrinsic meaning in relation to the underlying population, and hence are 'modifiable' and can be regrouped at any given spatial resolution. Wrigley (1995) summaries the scale problem as the tendency, within a system of modifiable areal units, for different statistical results to be obtained from the same set of data when that information is grouped at different levels of spatial resolution, whilst the zoning problem relates to the variability in statistical results obtained within a set of areal units as a function of the various ways those units can be grouped at a given scale. In practice, the scale and zoning problems interact, whilst the zoning problem is also influenced by spatial autocorrelation.

Classically the problem has been observed in the magnitude of correlation coefficients between variables, which increases as the size of the areas involved in the analysis increases (eg. Openshaw & Taylor, 1979). In terms of zoning, Openshaw & Taylor have demonstrated that, at any given scale, the zoning problem is likely to be sufficient to ensure that a wide range of statistical results are obtained, particularly with respect to correlations that have a tendency to change in magnitude and direction. More recently, the effects of the modifiable areal unit problem has been extended to include results of multivariate analysis, and it has been demonstrated that the goodness-of-fit statistic and parameter estimates are also sensitive to variations in scale and zoning systems (Fotheringham & Wong, 1991). In addition, an intrinsically related problem is the ecological fallacy. This arises when areal-unit data are the only source available to the researcher but the objects of study are individual-level characteristics and relationships. For example, when structural attributes for individual houses are only available as an average at the street level. In this case, relationships at a particular level of aggregation do not necessarily hold for the individual observations.

One solution to the modifiable areal unit problem, which has been implicit in much geographical work, is to assume that the problem does not exist. This was a common solution when spatial data were hard to come by, and there was little choice over the areal units. But many commentators have since argued (eg. Martin, 1996) that in the context of GIS, this should not the case. The advent of GIS has increasingly allowed access to finer resolution data and in a digital form that accommodates the design of various zoning systems. Openshaw (1995) argues that this has led to a user modifiable areal unit problem (UMAUP), since there is now greater freedom to produce different zoning systems and hence a range of different results. Thus there is a need to design zones that are intrinsically related to the objects of study if the results are to be meaningful. For instance, within hedonic research it has been argued that it is important to capture neighbourhood submarkets as precisely as possible, since failure to do this can lead to structural instability and incorrect estimates. Previously, administrative geographies were used to represent submarkets, due to a lack of data at other resolutions. However, with GIS and finer resolution data, the ability to aggregate the data into specified submarkets is greatly enhanced, and thus, so is the ability to generate a whole range of hedonic prices based upon different zoning schema. Therefore, to be meaningful, the zoning system needs to be constructed such that it accurately reflects the submarkets and this may rely upon some underlying concept or theory - see *Chapter Four*. By doing this, Openshaw (1995) argues that the modifiable areal unit problem will disappear, since it is only a problem whilst the influence of zoning systems are ignored.

3.5.3.2 Spatial Analysis and GIS

The second major function of a GIS is spatial analysis, which Goodchild (1987) has defined as the statistical description or explanation of either locational or attribute information, or both. From this description, it would seem that a GIS would be an effective tool for estimating hedonic house price models. However, in recent years there has been much debate concerning the use of GIS in spatial analysis. Although this is still ongoing, it has been acknowledged that GIS and spatial analysis are inextricably linked (Gatrell, 1991), although exactly what role GIS has is still unclear (Openshaw, 1994b). This uncertainty is summed up by Rogerson & Fotheringham (1994):

"Although GIS may not be absolutely necessary for spatial analysis, it can facilitate such analysis and may even provide insights that would otherwise be missed. It is possible, for example, that the representation of spatial data and model results within a GIS could lead to an improved understanding both of the attributes being examined and of the procedures used to examine them." (pp. 1-2)

This rather vague view of the role of GIS in spatial analysis can be accounted for by the fact that the main use of GIS has been one of data storage and management (Getis, 1994). Thus while GIS has been used extensively for storing, manipulating, transforming and visualising spatial data, its spatial analysis potential has been under utilised. Openshaw (1994a) argues that attention needs to be moved away from the view of GIS as solely one of Geographic Information Handling to one of Geographic Information Using. He continues this argument by providing ten basic rules for identifying GISable (sic) spatial analysis technologies. Three main spatial modelling issues are discussed within this framework. Firstly, the problems previously discussed of using data aggregated at various geographical scales, particularly those resulting from the modifiable areal unit problem. Secondly, he argues that the real secret of spatial analysis in GIS is that it should be able to handle special features of spatial information, rather than ignoring them. By 'special features', he is referring to the spatially structured nature of data precision and errors, since spatial data are rarely spatially random. Finally, he advocates GIS in assisting in exploratory data analysis, since he argues that users rarely know what patterns or relationships exist in the data, and that we are blind to the spatial patterns and processes that exits. In this way, the GIS can 'let the data speak for themselves' and suggest, with a minimum of pre-conditioning, what patterns might exist. (Openshaw, 1994b). The latter two aspects of spatial modelling shall be examined in more detail with respect to hedonic house price research.

3.5.3.3 Spatial Analysis Facilities Within GIS

There is a fairly substantial body of literature which concerns the interface between GIS and spatial analysis. These can be roughly spilt into two groups. Firstly, those concerned with the spatial summarization of the data, that is, the basic functions for the selective retrieval of spatial information within defined areas of interest, and the calculation of various summary statistics of this information. It is widely acknowledged that existing GIS offer a powerful array of techniques for spatial summarization, such as query facilities, Boolean operators, point-in-polygon and polygon overlay analysis, and various buffering techniques. Used alone, or in conjunction, these can identify and isolate specific geographic areas of interest and provide any relevant data concerned with the specified area. For instance, neighbourhoods with similar socio-economic and demographic profiles can be selected easily, whilst operations such as POINT-IN-POLYGON analysis can derive spatial information, such as average housing type in each neighbourhood, from the raw data held at different resolutions. However, Bailey (1994) asserts that such operations do not actually constitute 'spatial analysis', since they do not involve the analysis of patterns in spatial data, or the study of possible relationships between patterns and other attributes or features within the study region, or the modelling of such relationships for the purpose of understanding or prediction.

Much less has been written about the GIS operations with respect to such a definition of spatial analysis. Currently few GIS packages offer any capability of spatial modelling, in a statistical sense, of either raw or derived spatial data. Openshaw (1991; pp. 389) observes wryly that "[a] good GIS will today probably contain over 1000 commands ... but .. none will be concerned with what would correctly be termed spatial analysis rather than data manipulation". Gatrell (1994) comments that although this is a slight exaggeration, the essence of his argument is correct. Nevertheless, spatial analysis 'tool-kits' have been incorporated into GIS over the past few years. For instance, in ARC / INFO, techniques such as network analysis, routing, location/allocation modelling, and grid-based analysis has become standard spatial analysis tools in recent years. However, in terms of statistical spatial analysis, GIS is still very much in its infancy (Bailey, 1994).

Instead, most attention has been focused upon the links between GIS and spatial analysis. Goodchild (1991) has characterised the general types of links between GIS and spatial analysis in terms of 'fully integrated', 'tightly coupled' and 'loosely coupled'. A fully integrated linkage occurs when statistical spatial analysis software has been incorporated into the GIS package, although as it has just been discussed, this has been very limited and surprisingly slow. More attention has been focused upon 'loose coupling' and 'close coupling' of GIS and statistical software. In the former, data are exported and imported seemlessly into and from the GIS and statistical package whilst with the latter, the GIS allows custom written statistical functions to be embedded within the GIS. 'Loose coupling' GIS have been the most common form of linkage, with the coupling to an external statistical package or graphics software usually achieved through ASCII files exported from GIS.

It should be clear that although GIS is a valuable tool in spatial analysis, it has been under utilised. This has been mainly due to the neglect of spatial analysis in general, and statistical spatial analysis in particular, since both are in many ways fundamental to the effective use and exploitation of GIS in may different applied contexts (Openshaw, 1991). In *Chapter Two*, the problems of spatial structures inherent in spatial data, such as spatial dependence and spatial heterogeneity were discussed with respect to hedonic models. In this case, it is not only important to know the attributes of a house in a particular location, but also the relationship of these attributes to the attributes of houses in other locations. Getis (1994) calls this proximal space and has connections with spatial dependence. Since this spatial aspect of the data can be handled conveniently in a GIS, Sinton (1992) identifies this as the chief area for GIS research in the future: "I believe that the spatial inter-dependence among geographic entities is the theoretical linchpin of the GIS industry" (pp. 2-3).

In this sense, Bailey (1994) argues that the potential benefits of GIS are largely in facilitating the construction of proximity matrices between locations, known as W matrices, which are a necessary input to many the autocorrelation methods. Martin (1996) also notes that the fact that GIS is able to encode both location and attributes makes possible the development of techniques which incorporate explicitly spatial concepts such as adjacency, contiguity and distance that are important measures with respect to W matrices. Furthermore, the potential exists using the GIS to derive more sophisticated relationship measures between areal units which account for physical barriers, such as rivers, and network structures, such as roads. Such a technique would be beneficial within hedonic house price research since it has already been noted in *Chapter Two* that such models suffer from spatial autocorrelation caused by adjacency effects.

3.5.3.4 The Use of GIS in Previous Hedonic Research

The above discussion has illustrated the potential benefits of using a GIS in hedonic house price research. However, although GIS technology has been generally available for well over a decade, its use within hedonic house price studies has been rare, with the few recent exceptions including Waddell & Berry (1993), Waddell et al, (1993), Sanchez (1993), Cooley et al, (1995), Kennedy, et al (1996) and Lake (1996). The main use of GIS in these studies has been mainly to calculate distance and proximity measures in terms of both physical distance and time using tool-boxes such as *NETWORK* in ARC / INFO. Other uses of GIS has been to calculate lot-size from digitised boundary coverage's of landuse (Waddell & Berry, 1993), and the mapping of the error terms to determine the existence of spatial autocorrelation (Waddell et al, 1993). Hence, it would appear that GIS has been under represented in this field of research.

3.5.4 Conclusion

It is clear that GIS would be a valuable tool in the estimation of hedonic house price models. However, it is important to have a clear framework of the data environment if the GIS is to produce meaningful results. Also, it would seem that the main role of the GIS would be in the storage, manipulation, generation and visualisation of data, as opposed to the explicitly estimation of the hedonic price function. Thus, the GIS would be a loosely-coupled, with the data exported into statistical packages where the hedonic price function will be estimated.

Section 3.6 Conclusions

The aim of this chapter was to assess and evaluate the types of housing attributes that enter the house price determination process. In doing so, it has highlighted the problems associated with locational attributes. There has been very little consensus to the types of locational attributes that influence house price, and empirical evidence is contradictory. In particular, evidence supporting the existence of a negative rent gradient form the CBD outwards has been conflicting, and this has cast doubts upon the validity of the assumptions underpinning the micro-economic theory of housing markets. It would appear that locational attributes suffer from both conceptual problems and measurement problems. With respect to the former, traditional classifications of locational attributes based upon the concepts of relative and fixed location have been found to be problematic, with many attributes falling into both classes. This is acerbated by the measurement problems, with many locational attributes being historically poorly specified. This has caused violation problems in the estimation of the hedonic price function, such as multicollinearity and spatial effects. Together, these factors have contributed to inconsistent results and a general uncertainty to the influence of location upon house prices.

As a means of ameliorating these problems, to approaches have been identified. Firstly, the concept of locational attributes as being either relative or fixed has been replaced by the concept of locational attributes operating across different spatial levels. Such a concept allows locational attributes to behave more like externalities, with their influence determined by both magnitude of the attribute and proximity to the property. The second approach is to use a GIS as a median for the study. This will allow the measurement of locational attributes to be improved, and will also aid in the analysis and visualisation of the results. Therefore, *Chapter Four* is concerned with data collection, and shall discuss issues concerned with the availability and the accuracy of housing data. *Chapter Five* will then discuss the construction of a context-sensitive GIS, with a particular emphasis upon data integration and the generation of locational attribute data

Chapter Four
Cardiff Case Study and Data Sources

Section 4.1 Introduction

The previous chapters have provided the background to the research, and have raised the issues that need to be investigated. The aim of the first part this chapter is expand upon these issues in greater detail and to put the study into context. The remainder of the chapter will investigate the sources of data needed to fulfil these objectives, with particular emphasis upon the survey work undertaken to collect this information. Hence, section two and three provide an overview of the case study area of Cardiff and the aims and objectives of the study. Section four investigates the official and unofficial sources of data that are available for study of the built environment, whilst section five describes survey and fieldwork necessary to obtain some of this information. The final section concludes the chapter and discusses some of the aspects of the following chapter.

Section 4.2 The Cardiff Housing Market

4.2.1 Introduction

An important consideration is the delimitation of the areal units of study, in this case the housing market. It has already been discussed in the previous chapters that meaningful analysis demands the availability of the right sort of data for the right sort of units. Hence, results from the analysis of spatially heterogeneous process, such as housing market dynamics, can be influenced by the choice of areal units. However, as has been shown, the theory of how housing markets are structured is complex, and the examination of markets has frequently failed to address their definition, composition and structure (Adair et al, 1996). Most of the empirical work has shown a considerable degree of variation in the definition of housing market areas. The spatial definitions of markets have ranged from subcity areas, such as specific suburbs, to whole cities, local labour market areas and standard regions, depending upon the purpose of study and the availability of data. However, the boundaries of a housing market should be meaningful.

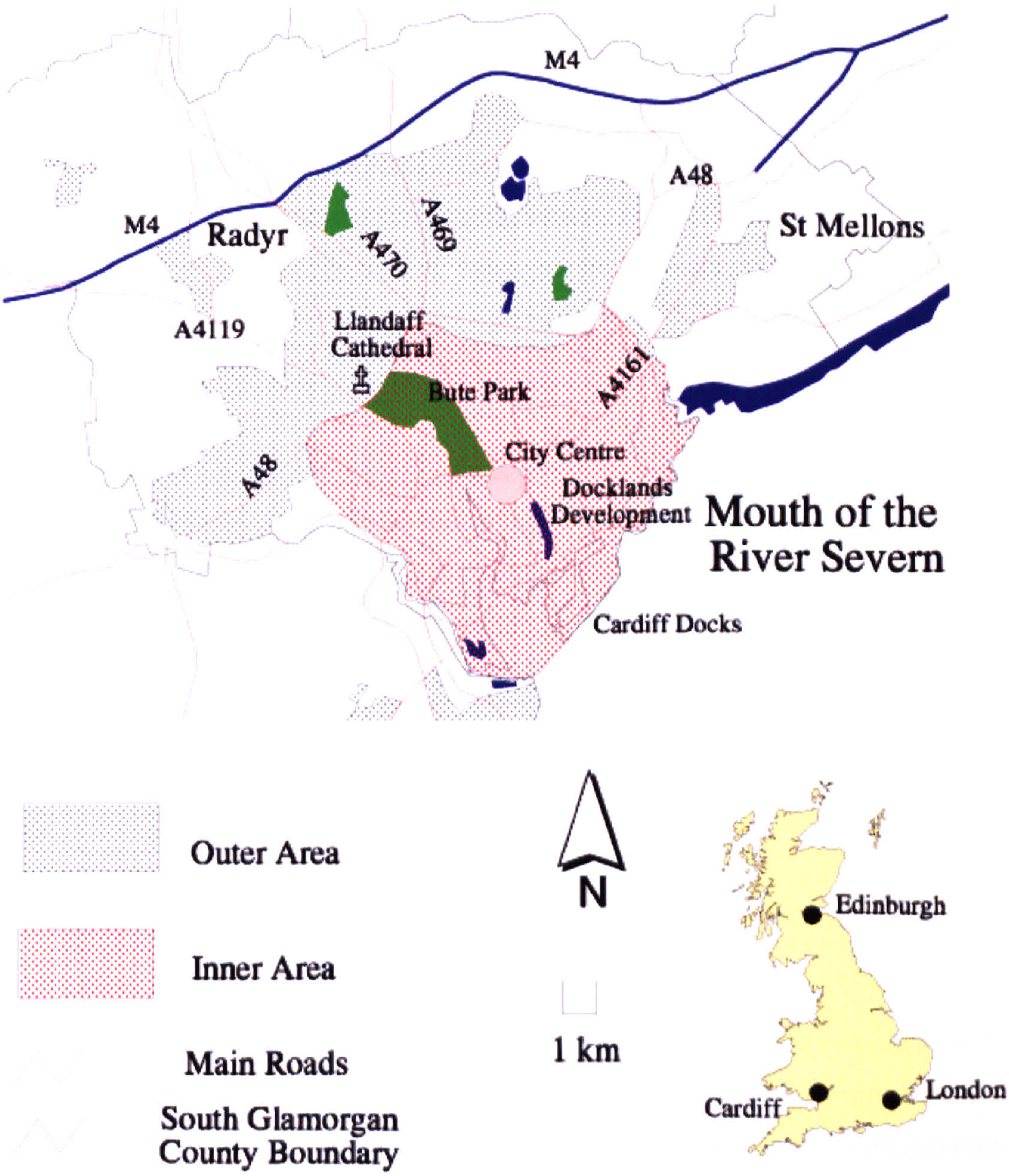
The chosen area of study in this research is Cardiff. Cardiff was selected because it had recently been used in work concerning the changing geographies of revenue raising (Martin et al., (1992); Longley et al., (1994)), the results of which have informed this study. The definition of the Cardiff housing market is the urban area bounded by the jurisdiction of Cardiff City Council. This contains both the inner city of Cardiff and its suburban hinterland. The actual boundaries are well defined, and they do not intersect any continuous built up areas. As such, the housing market can be regarded as autonomous and self-contained. A brief overview of Cardiff is described below.

4.2.2 Background History

Cardiff is the capital of Wales and was founded by the Romans, with Cardiff Castle in the centre of the city occupying the site of the original Roman fort. The Castle was added to in the nineteenth century, with the city rapidly developing as a port from the 1860s onwards, exporting coal from the South Wales Valleys to the rest of the world. Its development slowed markedly in the aftermath of the First World War when export trade slumped (Daunton, 1977). Growth during this fifty year period was very rapid, and as a result the inner city today still exhibits considerable homogeneity of built form. More recently this homogeneity has been disrupted to some degree by local authority estates and scattered pockets of infill redevelopment. Cardiff City Council has vigorously pursued a range of urban renewal programmes over the last couple of decades, and it is the implementation of such policies that the City Council has defined the 'Inner Area' to approximate this predominately nineteenth century urban core. The northern boundary of the Inner Area is defined by a major dual carriageway trunk road that physically splits the Inner Area from the suburbs. The eastern and southern boundary is formed by Cardiff Bay, whilst its western side is defined by the River Ely and green belt land. All together, the Inner Area covers approximately twenty four square kilometres. Beyond the Inner Area, the 'Outer Area' contains Cardiff's suburbs - see Figure 4.1. There are considerable differences between the Inner and Outer Areas in terms of dwelling age and type, and differences in household size and composition. As such, these two areas shall be discussed separately.

Figure 4.1

An Annotated Map of Cardiff



4.2.2.1 The Inner Area

The Inner Area is characterised by Victorian and Edwardian terraced housing, with early twentieth century semi-detached and detached private property located at its periphery. It has been identified by Cardiff City Council as a convenient unit for various aspects of urban policy, and remains a focus of sustained improvement and repair activity. There are a small number of Local Authority built housing estates, such as in Gabalfa and Tremorfa, and public sector purpose built flats in Butetown, although these do not represent a substantial proportion of the overall housing stock. The Inner Area also contains a small number of prestigious new housing projects, mainly associated with Cardiff Bay redevelopment scheme. It is also the location of many older multi-occupied properties sub-divided into flats and bedsits. The Area includes the commercial and financial centre of the city, and is also the location of cultural and recreational amenities such as the Castle, Cathays Park, the City Hall, Law Courts, the National Museum of Wales, and the University College.

4.2.2.2 The Outer Area

The suburbs contain lower density inter-war and post-war private housing and a number of extensive Local Authority estates. The western suburbs are characterised by huge post-war Local Authority housing estates of Ely and Caerau, and also the prestigious neighbourhood of Llandaff. The eastern suburbs are the location of inter- and post-war semi-detached houses, with modern flats and apartments in Pentwyn and the huge modern peripheral estate of St. Mellons. The northern suburbs are represented by the upmarket neighbourhood of Cyncoed, containing the majority of Cardiff's bungalows, with modern housing estates beyond.

4.2.3 Cardiff Housing Submarkets

A second, related issue, is the differentiation of the internal structure of the housing market. The principle of stratification of a housing market into subsets is widely recognised in the valuation literature (DeLisle, 1984) as a process of creating a number of homogeneous segments from a larger heterogeneous base. There are two broad approaches of stratification (Adair et al., 1996). The first is based upon the identification of distinct neighbourhoods which are readily distinguishable from one another primarily on the basis of environmental and locational characteristics. The second involves the identification of house groupings based upon differences in housing bundles, such as size, age and type. Both these approached were adopted, although it is only the former that is of interest in this section.

In the previous chapters it has been noted that the operational definitions of submarkets have been problematic. Most researchers have used census or administrative geographies, primarily due to issues of data availability. To be meaningful, these sub-divisions have to relate to small scale supply and demand mechanisms, and hence property attributes and household characteristics. As such, the physical delimitation of census areas have often tried to capture existing neighbourhoods, particularly since census geographies often following defined boundaries such as street, rivers and railway lines. These can act as barriers to the movement of the population, and tend to fix the boundaries of local social interaction (Knox, 1996). With respect to Cardiff, 'communities' are the basic areal administrative units, and these are also electoral wards. The term ‘community’ suggests some form of internal social cohesion, and it is a fact that Cardiff estate agents use these communities as a basis for defining residential neighbourhoods - see section 4.5. Therefore it seemed sensible to use the same boundaries as an operational definition of submarkets for the whole of the housing market, resulting in twenty six submarkets in all - Figure 4.2.

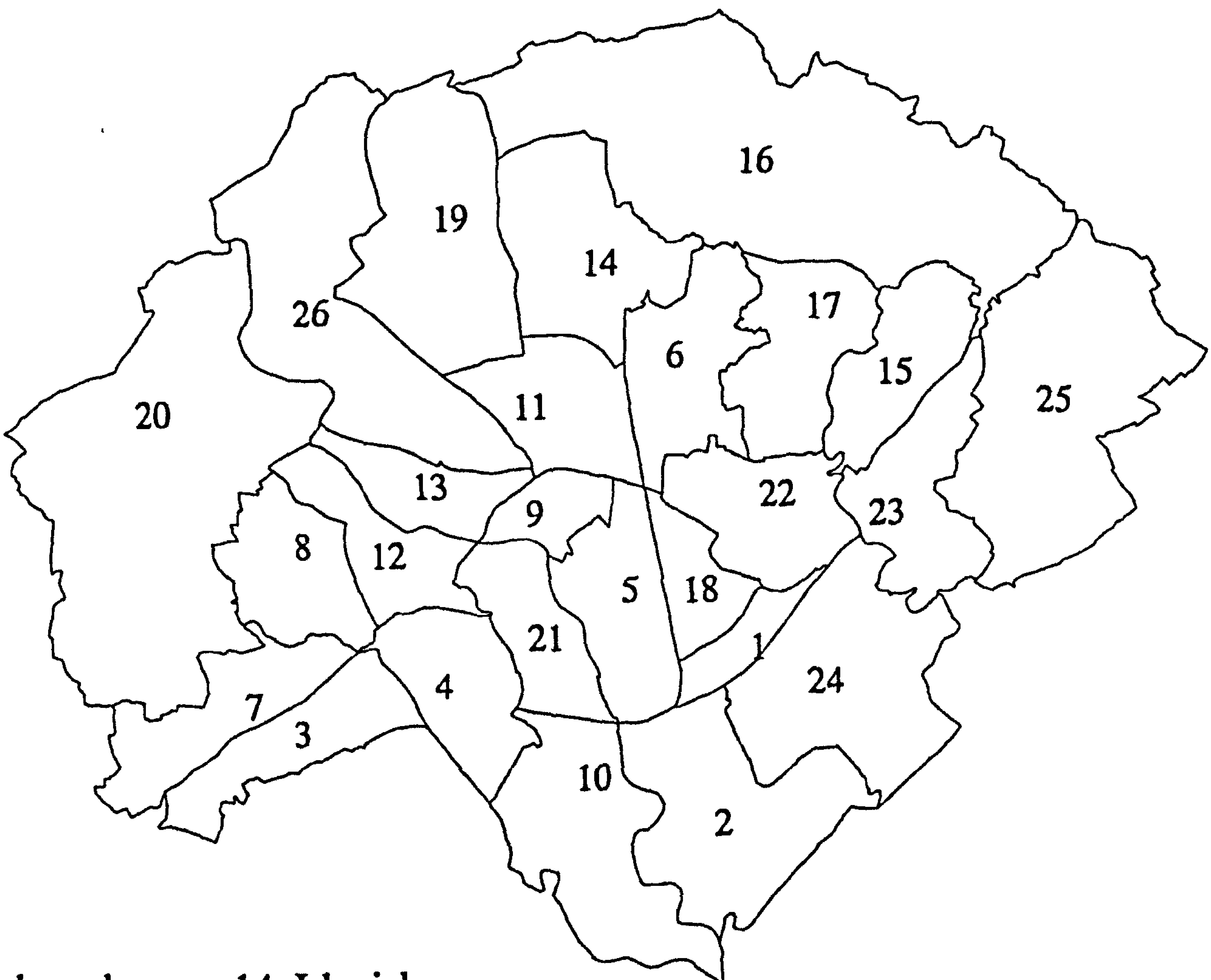
Furthermore, due to the obvious differences in housing stock and household composition, and its autonomous nature, it was decided that the Inner Area could be analysed separately, since the supply and demand schedules would be distinct from the rest of the city. The Inner Area is also an interesting area to study the effects of location since it is more heterogeneous than suburban locations, with property prices varying notably across smaller areas. Hence, it could be argued that locational externalities play a more significant part in price determination than in the Outer Area of Cardiff.

The Inner Area contains nine complete communities, and the two partial communities of Roath and Llandaff. These in turn can be completely sub-divided into the eighty one Housing Condition Survey (HCS) areas. These were defined by Cardiff City Council on the basis of within-area homogeneity of built form and residential characteristics, in order to facilitate the detailed implementation of its housing policy.

Therefore, the housing market has been divided along two criteria. Firstly, the whole housing market has been stratified into twenty six submarkets based upon communities which local estate agents use as a rough definition of neighbourhood. Secondly, the housing

Figure 4.2

Map of the Twenty Six Communities



1 Adamsdown
2 Butetown
3 Caerau
4 Canton
5 Cathays
6 Cyncoed
7 Ely
8 Fairwater
9 Gabalfa
10 Grangetown
11 Heath
12 Llandaff
13 Llandaff North
14 Llanishen
15 Llanrumney
16 Lisvane & St Mellons
17 Pentwyn
18 Plasnewydd
19 Rhiwbina
20 Radyr & St Fagans
21 Riverside
22 Roath
23 Rumney
24 Splott
25 Trowbridge
26 Whitchurch & Tongwynlais

market can be divided into an Inner and Outer area, with the Inner Area corresponding to the nineteenth century urban core. This, in turn, has been further sub-divided into eighty one HCS areas, based upon homogeneity of dwelling stock and resident population. These sub-divisions of the Cardiff housing market form the basis of the study.

Section 4.3 Aims and Objectives

The first three chapters of this research has discussed the problems of valuing the built environment in detail. This has included the problems of gaining access to detailed, locationally referenced, disaggregated data, of how such data can be manipulated and the difficulties in modelling spatially referenced data using the hedonic price specification. Now these problems have been discussed, this section will formally identify the research aims.

The first objective is to identify potential data sources that will allow the built environment to be valued, and in particular, locational externalities. It has previously been discussed that past research has relied upon data that has been poor with respect to its spatial resolution. This has caused problems with the subsequent modelling, especially with respect to spatial effects. Hence, an aim of the research is to acquire data that are referenced to a high resolution, specifically ~~the~~ at level of the individual property. These data will have to incorporate both the physical structure of the built environment, such as structural attributes of housing and the location of amenities such as shops and parks, and the social structure, such as household characteristics. Furthermore, data sources pertaining to a form of valuation of the built environment will need to be acquired. This is discussed in the remainder of the chapter.

The second objective is to create a Geographic Information System (GIS) to assist with the research. It was previously discussed in *Chapter Three* that a GIS is an ideal medium for such research. It is capable of storing and manipulating physical and socio-economic data at a variety of spatial scales, and can aid spatial analysis. Despite this, there has been very little use of GIS in past hedonic house price studies. Section 4.2 has described how the Cardiff housing market can be divided into an Inner Area and an Outer Area to allow separate analysis of the former. This means that two separate GIS's will need to be created: one for the whole housing market and one for the Inner Area. Moreover, since the Inner Area will be the focus of much more detailed research, it is this GIS that will require the most work. This is discussed in *Chapter Five.*

The third objective is to use the GIS to generate measures of locational attributes. As was discussed in length in *Chapter Three*, locational attributes have been historically poorly specified, and this has been the result of a lack of good data. The GIS should to some extent ameliorate this problem since it is capable of manipulating both physical and social data at a variety of resolutions, and deriving new spatially referenced data very efficiently. This is also discussed in length in *Chapter's Five & Six.*

The fourth objective is to model the spatial dynamics of the Cardiff housing market. This will involve exploring how housing attribute data can be modelled efficiently using the three hedonic specifications that were described in *Chapter Two*. More specifically, *Chapter Seven* will investigate the ability of the following specifications to model spatial data:

I. The traditional hedonic specification
II. The spatial parameter drift specification
III.The multi-level specification

The corollary of this will be an examination of how structural and locational attributes interact within the housing market, and how this affects spatial price differentials.

The fifth, related objective, is an attempt to model the Cardiff rent gradient. This is a basic tenant of urban economics, but its existence has proved to be elusive and controversial. It is hypothesized that this is due to inferior spatial data, and poor modelling techniques, as opposed to theoretical concerns. Once these have been ameliorated, the rent gradient for Cardiff should be identifiable.

The sixth objective to explore in detail how locational externalities operate within the housing market. This is described in *Chapter Eight*, which investigates how locational externalities are incorporated into the housing market mechanisms of the Inner Area. The research is particularly interested to discover the level of resolution that locational externalities operate, and the implications that this has for the property prices and the built environment in general.

Hence, the above six objectives form the basis of the remainder of the research. The objectives fit within the context of previous studies, although the research aims to expand these by the use of alternative specifications of the hedonic house price function, and the use of a GIS as a tool for spatial analysis. An important aspect of the research are the two different case studies. The first uses data for the whole of Cardiff in an attempt to model the spatial dynamics of the housing market. As such, this examines the housing market in more general terms, and the use of locational attributes are limited. The second cases study focuses in on the Inner Area, using the findings of the previous study to model locational externalities in finer detail. Hence, the two studies complement each other, with the former providing the basis for the latter. The remainder of this chapter investigates potential sources of data available for valuing the built environment.

Section 4.4 Sources of Property Related Data

4.4.1 Introduction

The research has utilised a whole range of large and complex socio-economic datasets at various levels of disaggregation. Although there is very often a difficulty in obtaining and managing locationally disaggregate data on eligible individuals and their residences, several data sources have been identified. These include both primary, field work sources and secondary data sources. Three of them - the rates register, the council tax register and the Cardiff Housing Condition Survey - have been utilised in previous research in relation to the changing geographies of revenue raising within the Inner Area of Cardiff (Martin et al, 1992; Longley et al 1993; 1994). However, a substantial part of this chapter will refer to the appraisal and collection of new data sources pertaining to property valuations, property attributes and environmental assessments. Hence, the first part of this section will briefly evaluate the different sources of property valuation data. Other sources of data available for the local housing market shall then be discussed, with particular emphasis upon selling price data. Finally, the section shall conclude with an evaluation of the Cardiff Housing Condition Survey and the 1991 Census as a source of data for property attributes and household characteristics.

4.4.2 Property Values

The independent variable in an hedonic model is usually a measure of the valuation of a property. It is possible to have a whole range of valuations for a single property at a particular moment in time, depending upon the purpose of the valuation (Millington, 1990). However, this research shall only be considering two types of property values; use value and exchange value. Use value generally refers to the net utility supplied by the bundle of housing services. Although use value is not fixed by the attributes of a property alone, since the utility gained from these services will vary between individuals and households, the measurement of use value has been most commonly equated to the rental value of a property in a rent-clearing market (Harvey, 1973). As discussed below, in a UK context, such a measurement can also be analogous to the rateable value of a property. The exchange value of a property is the capital value it can realise in a competitive housing market. Although the use value of a property is a major determinant of its exchange value, this will also be influenced by the property's potential for increasing capital gain since, as was explained in *Chapter One*, housing is also a major source of stored wealth. In terms of exchange value, this research is interested in selling price and council tax valuations. Hence three types of property valuation data will be investigated; rateable value, council tax band and selling price. Since sources relating to the former have previously been investigated as part of the research into the changing geographies of revenue raising in Cardiff, they shall only be briefly summarised here.

4.4.2.1 Rateable Values

The rateable value of a property is assessed as the net rental value of the property (Foster et al. 1980; pp. 309). However, such an assessment has been increasingly notional due to the virtual disappearance of the private rented housing sector in the UK. Nevertheless, with some assertions, rateable value can be equated to the use value of a property, particularly with respect to a property's structural attributes. The basis of the property rating data is the rates register, which is publicly available. In the case of Cardiff, this has been continuously updated but does not include properties built since the inauguration of the community charge in April 1990. Each rating hereditament has a unique property reference number (UPRN) which incorporates its street and community identifiers, a brief description of the property, an address and a rateable value - see *Chapter Five*. A rating hereditament is the portion of a building, a building or group of buildings treated as a unit for tax purposes.

Cardiff City Council has supplied the most up-to-date (in 1991) domestic rates register in machine readable form (see Martin et al, 1992).

4.4.2.2 Capital Values

In the majority of hedonic studies, capital value has been equated to selling price, although selling price and capital value need not be identical since buyers and sellers may be unaware of the true capital value or agree on a lower price for reasons of speed, convenience or family agreements- see *Chapter One.* Although the majority of this research will concentrate upon selling price as the measure of capital value, council tax valuations will also be considered. Council tax valuations place a property into one of eight taxation bands (A-H), based upon the selling price of the property in April 1993. The problems associated with the council tax have previously been discussed in Longley et al., (1993; 1994), and shall not be discussed in detail here, although it should be mentioned that the bandings are very coarse and are arguably not very accurate, particularly in areas of heterogeneous housing stock. The source of council tax data is the council tax valuation register. A draft copy of this register for the Inner Area of Cardiff was made available by the Valuations Office Agency. The council tax register contains a record for every residential property, including a brief description of the property type, the council tax band and a UPRN. The rates register represents the historical origins of the UPRNs used in the council tax register, although the two differ in practice as a result of a number of ad hoc procedures by made the Valuation Office Agency (see Longley et al, 1994).

4.4.2.3 Sources of House Price Data

The principal property value used in this research is selling price. However, a key concern relates to access to this information, and this problem is two-fold (Estates Gazette, 1985). Firstly, due to legislative restrictions, information pertaining to sales and valuations of individual properties by financial institutions are in the main confidential. Secondly, there is an absence of a central register of all sales. Only the Inland Revenue Valuation Office has full knowledge of all property dealings in England and Wales (Dixon, 1992), but this is not publicly available. In comparison, both the rates register and the council tax register are in the public domain, although access to the latter is more restricted. Hence, it is clear that the status on access to property valuation data is blurred. Commentators such as Dixon (1992) and Wyatt (1994) regard this mismatch as the key problem facing the property industry in

England and Wales to date. Therefore, this section will review and critique the sources of data available on the selling price and attributes of property. The emphasis will be specifically upon the local housing market, and the types of bias that may be encountered when using different sources of data. The section will conclude with a review of a recent housing condition survey of Cardiff, and its potential as a data source in this research.

4.4.3 Official and Unofficial Data Sources

House prices and their associated data are available in varying scope and detail from a number of different sources. Fleming and Nellis (1981) have made the distinction between official sources, where prepared data are derived from mortgage lending institutions, and unofficial sources, where unprepared data are obtained from other agencies that deal with property transactions. Data from official sources are typically complied by the government or by individual building societies and are available in prepared formats. At present there are at least fifteen such sources (Fleming & Nellis, 1992), a complete collection of which are detailed in *Spon's House Price Data Book* (Fleming & Nellis, 1987). Unofficial sources include institutions such as estate agents, local authorities and surveyors, where the data are not usually available in a pre-prepared format. These two sets of sources have been reviewed for their potential suitability as a source of house price data.

4.4.3.1 Official Data Sources

There are currently four principal official sources of house price data in the UK: two prepared by the Department of Environment (DoE), and data prepared by the Halifax and the Nationwide building societies (Nicol, 1996). The underlying motivation for the collection of such data is the derivation of standardized indices to measure house price change over time. However, the information available varies, particularly with respect to the structural and locational attributes of a property. Also, as will be explained, each of the four sources differ in their propensity for bias. Hence, below is a brief review of their value as a source of data for research at the local housing market level.

I. The Department of the Environment Surveys

The main DoE source is the DoE / Building Society Association survey; the BS4 survey. This survey obtains information from a national sample of mortgage completion's, based upon returns from a sample of building societies and also the Abbey National Bank (Nicol, 1996). Whilst this provides the most complete picture of national house prices, it can only provide average prices on a national basis, and does not present detailed, disaggregated information for local housing markets. For a more detailed review of house prices, the DoE makes available a five percent sample of the BS4 survey. This sample is based upon an average of around 25,000 properties per year since 1990, and since 1992 has also included information from non-building society lenders (Fleming & Nellis, 1994). Although the data are disaggregated to a greater extent than the BS4 survey, it is limited to four key variables: age, dwelling type, number of habitable rooms and regional location of the dwelling.

II. The Building Society Surveys.

The Halifax[1] and the Nationwide are the two largest building society mortgage lenders in the UK, with the Halifax alone having a 25% share of the building society market (Fleming and Nellis, 1994). Both possess an extensive database, although the Halifax is roughly three times as large as the Nationwide's having an average of 12,000 house transactions per month compared to the Nationwide's 4,000 transactions (Nicol, 1996). Unlike the DoE surveys, both data are recorded at the mortgage approval as opposed to the completion stage. However, they are over twice as large as the five percent sample, and are disaggregated to a far greater degree. Both databases contain information relating to the structural attributes of the property, but the Nationwide's data are far more locationally sensitive, with each property fully postcoded.

4.4.3.2 Problems and Criticisms of Official Sources

Fleming and Nellis (1985) have argued that the representativeness of the data provided by individual building societies needs to be interpreted with three factors in mind: geographical coverage, lending policies, and sample sizes of different societies. The first problem concerns the fact that many building societies are regional, or tend to concentrate their business in specific areas of the country. This factor is not a problem with the DoE survey's, since they sample a range of building societies, but can become an issue with data obtained solely from the Halifax or the Nationwide, especially with respect to local housing markets (Nicol, 1996). A related problem is the issue of the lending policies of different societies. A

[1] The Halifax has since changed its status and is now The Halifax plc

common criticism levelled against the four sources is that, since they are principally based upon transactions financed solely by building societies, they are subject to bias as they do not include transactions financed either by other institutions or cash sales. For instance, as a result of the removal of regulatory controls during the early 1980s, banks increased their share of mortgage finance from 8% in 1980 to 21% by 1989 (Beaverstock et al, 1992), representing a significant stake in the market. It has also been suggested that certain financial institutions have refused to deal with specific property types or within certain areas. This is known as redlining (Boddy, 1980) and can cause sampling bias if lenders concentrate on particular segments of the market. For instance, the tendency for banks to concentrate at the higher end of the property market implies that data taken solely from building societies may lead to a bias against more expensive properties (Fleming and Nellis, 1985), whilst some lenders are biased against council house sales. Although Fleming and Nellis (1994) have argued that this factor has been overstated, the problems may still be relevant for house prices at the local level. The final factor, sample size, may be exacerbated by both the issues of geographical coverage and the related problem of lending policies. Data obtained from a single building society may be very selective and biased at the level of the local housing market, with specific property types and neighbourhoods under-represented.

Nicol (1996) has identified two additional problems when using data obtained from official sources. Firstly, since the two building societies record data at the mortgage approval stage, Nicol argues that recorded house prices may be reduced, or sales breakdown, before mortgage completion as a result of surveyors reports or housing chains breaking. Fleming and Nellis (1985) have found that this occurs infrequently, although Nicol has argued that there is a danger that certain neighbourhoods or house types or types of purchaser may have greater propensity to 'vanish' between approval and completion stage, and this seems to be particularly true for first time buyers.

The second problem relates to data validation and editing procedures (Fleming & Nellis, 1994). Data in all four sources are cleaned before being added to the database. Data that are eliminated include data-recording errors, properties that are sold at non-market prices, such as council house sales and those sold to relatives or sitting tenants, and any other spurious returns. The degree of cleaning varies between the four sources with the Nationwide employing tighter criteria, followed by the Halifax and the DoE. Although cleaning the data is necessary, it does increase the danger of bias within the official sources, particularly with the removal of certain property types, and properties in certain neighbourhoods.

4.4.3.3 The Nationwide Database and the Cardiff Housing Market

An important point to consider when using official data sources is its applicability to a small scale study of a local housing market. Fleming and Nellis (1992, 1994) extensively review the extent and range of data available, and with the exception of the Nationwide database, it would appear that data from official sources are disaggregated only at a coarse resolution, specifically at the scale of the standard region. These data are therefore useless for studying local housing markets, which is partly due to confidentiality considerations. Data from the Nationwide Building Society are classified at three geographical levels: region, neighbourhood and surrounding area. Moreover, the postcode of each property is also recorded. Hence, in terms of research at the local scale, the Nationwide Building Society data are the most appropriate.

To this end, the Nationwide Database of mortgage approvals was acquired for 1994 and 1995. This amounted to roughly 40 000 transactions for each year for the UK. With respect to Cardiff, the Nationwide Building Society approved 442 mortgages in 1994 and 387 in 1995. The database includes a large number of variables, including valuation (asking) price, purchase (selling) price at the mortgage approval stage, the date of purchase, and the type of property. These data were extracted for Cardiff and used as a reference for the data collected from the unofficial sources. This shall now be discussed.

4.4.3.4 Unofficial Data Sources

Unlike official data sources, these include all of the organisations that have dealings with housing transactions, and not just mortgage lending institutions. These are typically represented by the estate agencies that co-ordinate the buying and selling of property, although in recent years this estate agent work has also been undertaken by banks and building societies as well as a number of insurance companies. This has occurred due to acquisitions of traditional estate agents by such financial organisations during the 1980s. Between 1983 and 1988, thousands of estate agencies were absorbed into corporate structures throughout England and Wales, resulting in a polarised structure of a limited number of large firms and large number of small firms (Beaverstock et al, 1992). Figure 4.3

reflects this fact, suggesting that on a national basis in 1990, over 28% of housing transactions were co-ordinated through just eight large firms, with nearly half of the

Figure 4.3

Percentage of UK Housing Transactions, First Half of 1990

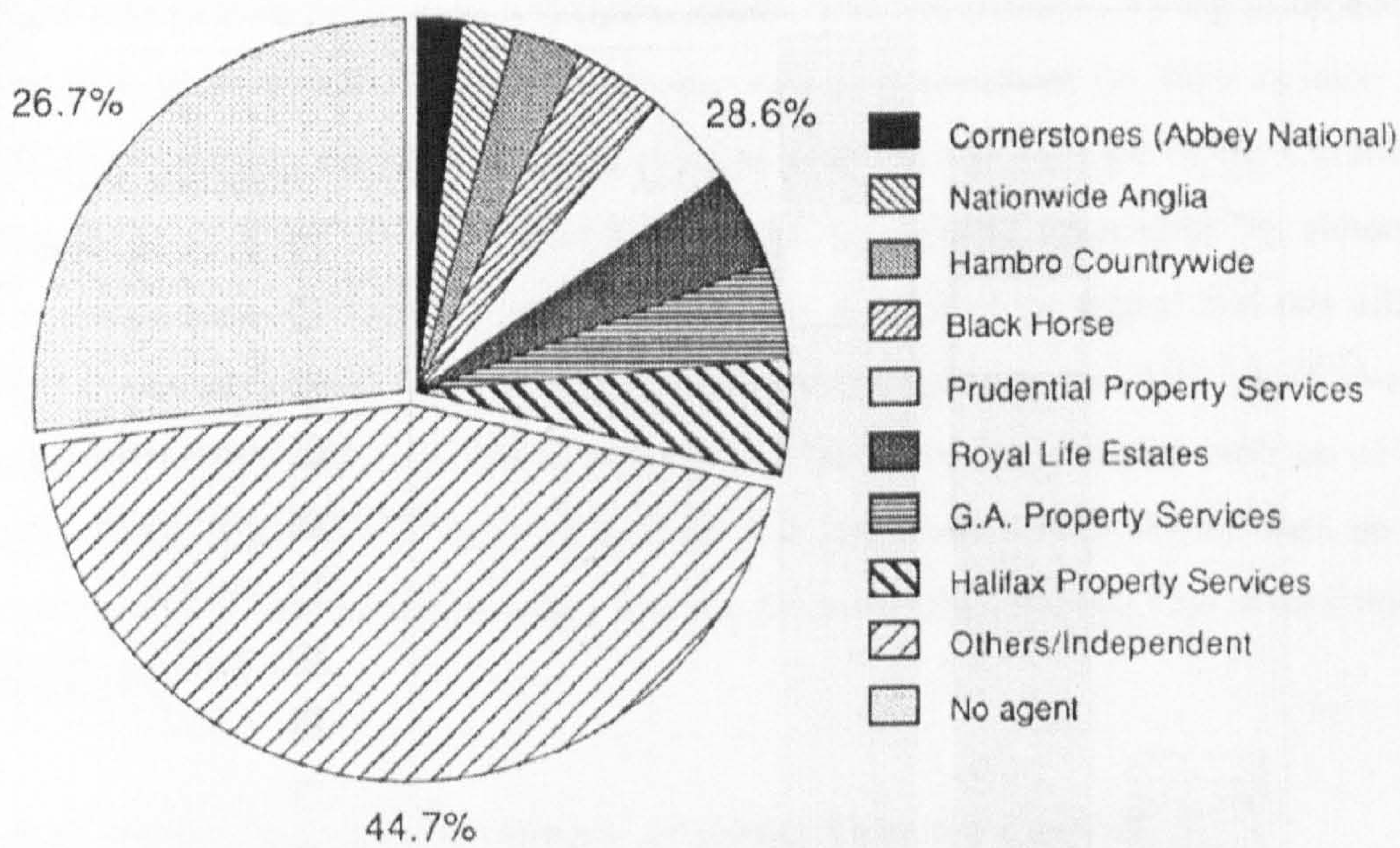


FIGURE 1. Percentage share of UK housing transactions, first half of 1990
Source Based on data in J R Adams & Associates, Adams Residential Property Index, 1990, quoted in Estate Agency News, Sept. 1990

Source: Beaverstock et al., (1992) pp. 173.

remainder occurring through independent estate agents or local financial institutions, and the rest by the individual household. Since the collapse of the housing market in the late 1980s, many of the large financial companies have re-sold their branches back to the smaller, independent estate agencies. However, estate agent operations are likely to remain a part of most large financial institutions (Goodman, 1990) and hence will still represent a significant share of the market.

4.4.3.5 Problems and Criticisms of Unofficial Sources

The advantage of unofficial sources of data is that the property's full address is obtainable, as is as a wide range of structural attributes. Also, because a whole range of different institutions that deal with housing transactions can be sampled, the problems of sampling bias as discussed previously can be reduced. However, there are two main problems with unofficial sources of house price data. Firstly, by its nature, the data are not in prepared formats, and the type of information available will vary between sources. Secondly, and more importantly, the actual information on house price will be restricted to asking price due to confidentiality considerations (Freeman & Dixon, 1992). This is the principal difference between official and unofficial data sources, since official sources record in effect selling price. Asking price represents the market valuation, usually as assigned by a professional estate agent, and is typically higher than the eventual selling price due to the 'valuation gap'. Cardiff estate agents were asked to comment on their opinion of the difference between asking and selling price of properties at the time of the Cardiff house price survey. Figures varied between 4 - 9%, with the majority suggesting 7%, although this was said to vary between area and property type. It can also be argued that this difference may vary systematically between agencies, since some agencies may deliberately over-value properties to gain customers due to competition. However, the potential problem of bias by using asking price can be lessened is if used in conjunction with official data on selling price from the Nationwide Building Society for same time period. This is described in a later section.

4.4.4 Other Sources of Property Attribute Data for Cardiff

The remainder of this section shall examine the two other sources of property related data for the Cardiff housing market. These are the Cardiff Housing Condition Survey and the 1991 Census, both of which have data relating to the structural and locational attributes of property, as well as household characteristics, although they vary in both detail and level of disaggregation.

4.4.4.1 The Cardiff Housing Condition Survey

The Cardiff Housing Condition Survey (CHCS) was commissioned by Cardiff City Council in November 1988, undertaken in the first half of 1989, and reported in November 1989 (see Keltics, 1989). It provides a detailed picture of housing and environmental conditions, as well as the socio-economic characteristics of occupying households, of the private sector housing stock within the Inner Area of Cardiff. The research was based upon an interval sample survey of 1 in 5 consecutive domestic dwellings within the Inner Area which were owned by either the occupiers, private landlords or housing associations. The sampling frame used was based upon dwelling in the rates register, and in total, a sample of 7,413 dwellings were drawn from a screened total of 37,115 privately owned or housing

Table 4.1
Summary of Data Available in the CHCS

Physical Survey	Social Survey
Survey reference number	Persons in household
Total repair costs	Sex of head of household
Dwelling type	Age of head of household
Number of rooms	Employment status of head
Property vacancy	Occupation of head of household
Date of construction	Length of residence of head
Representative of type	Previous ownership staus
Number of habitable floors	Length at previous address
Number of dwellings in flat	Sex of spouse
Number of floors in flat	Age of spouse
Non-domestic use in building	Employment status of spouse
Prescence of central heating	Owned or rented
Standard of heating	Furnished or unfurnished
Dwelling house multi-occupancy	Type of tenancy
Number of persons in building	Sharing WC
Number of separate lettings	Sharing bathroom
Habitable rooms adequately heated	Quality of local shops
Adequate property management	Quality of public transport
Garage	Quality of access to city centre
External WC	Quality of sports facilities
Front garden length	Quality of local parks
Rear garden length	Quality of play spaces
Condition of roads / pavements	Quality of community facilities
Quality of street environment	Length of residence expected
Condition of rear lanes	Household access to car / how many
Visually obstructive non-residential landuse	Quality of physical environment
Atmospheric obstructive non-residential landuse	Traffic problem
Noise obstructive non-residential landuse	In street: standard of upkeep
Non-residential use in dwelling	In street: presence of eyesores
Derelict land / build in vicinity	

association owned dwellings. To aid the research, the Inner Area was disaggregated into eighty one housing condition survey (HCS) areas, which were chosen by the City Council for their high degree of internal homogeneity of the dwelling stock. The HCS areas aggregate precisely into communities, and the mean size of each area is approximately 450 dwellings. The survey was designed and implemented by two organisations; Keltecs (Consulting Architects and Engineers) Ltd, who were responsible for the physical surveys of the dwellings, and Research and Marketing (Wales and West) Ltd, who were responsible for the socio-economic surveys of the residents. To date, it is the most detailed, locationally disaggregated private sector house condition survey ever carried out in British city.

I. Survey Questions

The basis of the survey questions was the English Housing Condition Survey, although these were tailored for the specific needs of the Cardiff housing stock and resident population. Each record in the CHCS contains data from both the physical and social surveys, as well as a UPRN and a HCS identification number. The UPRNs in the CHCS were based upon the UPRNs in the rates register, from which the sampling frame had been developed. Table 4.1 is a summary of the data in the CHCS. The physical survey was divided into two distinct parts: an external survey and an internal survey. Both surveys were concerned with the structural attributes and condition of the property, with particular emphasis upon state of repair. In addition, an assessment of the environmental quality was also undertaken, with emphasis upon non-residential land-use. The social survey covered questions that can be split into the following categories; household membership, ethnic origin and financial circumstances; tenure of present occupiers; housing history and aspirations; degree of satisfaction with present housing; repair and maintenance of property; attitudes towards the local environment; applications to the local authority for house renovation grant aid; vehicle ownership and parking facilities; perceptions of housing maintenance within the neighbourhood and environmental pollution and nuisance problems. The data from both physical survey and the corresponding social survey were used to calculate an estimate a dwelling repair cost for each property using a repair cost model (Keltecs, 1989).

II. Survey Problems and Response Rate

Table 4.2 is a summary of the responses rates for both the physical and social surveys. The overall response rate to the social survey was 60.5%. Generally, the response rate was in the range 55 - 65% in each ward, although this varied from 51% in Butetown and Llandaff to 69% in Splott. Similar patterns of variation in response rate are indicated for both the internal survey, and for the collection of sufficient data in both surveys to allow the calculation of a dwelling repair cost. The lower than anticipated level of contact, together with dropout or persistent non-contact after interview, posed a considerable problem. There was particular hostility to financial questions although there was no significant resistance to other questions, including dealing with ethnicity. One problem that could not be overcome in sampling was the way Houses in Multiple Occupation (HMOs) appeared on the rating list. Some HMOs were rated as a single unit and were sampled in the normal way. However, other HMOs contained separately rated units and may therefore have been over represented because of multiple entries on the sample frame (Keltecs, 1898; pp. 27). The problems of HMOs, and their representation in different administrative datasets is a pervasive one, and is discussed in more detail in the next chapter.

Table 4.2
CHCS Response Rate by Community

Community	Dwellings on Rates Register	Dwellings on Sample Frame	Completed Social Survey	Completed Physical Survey
Butetown	448	90	46 (51.1%)	37 (41.1%)
Cathays	4437	857	522 (60.9%)	489 (57.1%)
Gabalfa	1941	388	261 (67.3%)	233 (60.1%)
Roath (part)	2632	525	344 (65.5%)	301 (57.3%)
Plasnewydd	6382	1274	764 (60.0%)	687 (53.9%)
Splott	3308	661	456 (69.0%)	429 (64.9%)
Adamsdown	2956	591	348 (58.9%)	344 (58.2%)
Canton	5388	1078	603 (55.9%)	524 (48.6%)
Riverside	4879	974	537 (55.1%)	500 (51.3%)
Llandaff (part)	508	101	52 (51.5%)	44 (43.6%)
Grangetown	4236	884	553 (65.5%)	515 (61.0%)
Total for Inner Area	37115	7413 (20%)	4486 (60.5%)	4104 (55.4%)

Source: Keltics (1989) Appendix 3 Table I

4.4.4.2 Census Data

Table 4.3 is a summary of the variables constructed from the 1991 small area statistic Census data at Enumeration District and Community (Ward) level, and reflects the aim of constructing indices of deprivation and socio-economic classification for the whole of the Cardiff housing market. This follows similar methodologies used in many hedonic house price studies for constructing surrogates for locational attributes, and was necessary because data in the CHCS relating to the physical environment and social characteristics were only available for the Inner Area. The choice of variables used to construct these indices are important, since they can determine the results of the analysis. Hence, the variables were chosen to represent the factors considered to influence housing supply and demand and thus residential differentiation (e.g. Hirschfield et al, 1995; Blake & Openshaw, 1996), and can be grouped into three general categories; socio-economic, family status/life-cycle and ethnicity; although it should be noted that these need not be mutually exclusive.

Table 4.3.

A Selection of 1991 Census Variables

Socio-economic dimension	Percentage of male unemployment Percentage of female unemployment Percentage of lone parent households Percentage of households with no car Percentage of households with two or more cars Percentage of households with shared bath / shower Percentage of households with no bath / shower Percentage of households with shared inside WC Percentage of households with no inside WC Percentage of households with no central heating Percentage of households in owner occupied tenure Percentage of households in Local Authority tenure
Family life-cycle dimension	Percentage of households young and single Percentage of households pensioners Percentage of households married with family
Ethnic dimension	Percentage of non-white households

The socio-economic dimension is perhaps the most important category since this can provide proxy information relating to general affluence and deprivation. The first five variables relate to individuals within households, and can be roughly divided into those indicating areas of relative deprivation (the first four variables), and areas of relative affluence (the fifth variable). For instance, because of the high running costs, car availability, and especially the availability of two or more cars, has been used by many researchers as a measure of short-term deprivation (e.g. Townsend et al. 1986). Similarly, because of their social needs, lone parent households have been highlighted as an important group to differentiate (Blake & Openshaw, 1996). The final seven variables in the socio-economic category relate to the different forms of accommodation and amenities. These are important since they can be used as a measure of lack of resources and residential security, particularly with respect to tenure types. For instance, due to the financial commitments, owner occupation is seen as a surrogate for long term financial stability, in contrast to accommodation rented from Local Authorities. The lack of basic amenities are used as a measure of deprivation, with the lack of central heating being a more recent addition as the proportion of dwelling which lack baths and WC's decline.

The family life-cycle, demographic dimension is important since lifestyle groups tend to concentrate in certain areas and can reflect differences in affluence and deprivation. In particular, households with young, single people and pensioners tend to be related to areas of low incomes and deprived areas, whilst the opposite is true for households with families. Finally, the ethnic dimension is important since the urban housing market can be structured along racial lines, although as previously discussed, this phenomenon is less prominent in a British context. Also, there is a tendency for many multi-ethnic areas to be associated with poorer areas, although there is a danger of overstating this relationship.

Section 4.5 The Cardiff House Price Survey

4.5.1 Introduction

It has just been discussed in section 4.4 that official sources of house price data are not generally applicable to local scale studies of housing market due to the lack of disaggregated data at this level. Instead, it was suggested that data from unofficial sources, namely estate gents, should be used since these are available at the resolution of the individual property. Hence, this section describes that methods used to obtain such information, bearing in mind the problems of sampling bias that were described in the previous section.

As Figure 4.3 suggests, in 1990, nearly half of all housing transactions in the UK occurred through independent estate agents or local financial institutions whilst the remainder were dealt with either by the eight largest corporations (28.6%) or by the individual householder (26.7%). Although this situation will vary across the country and is likely to have changed since 1990, it is probable that each of the three groups still represent significant sources of transactions in the housing market. Hence, all three groups will have to be surveyed to prevent sampling bias. The geographical area covered by each agency also will have to be considered. Estate agencies have traditionally been small, and although the coverage of one particular agency may represent the entire local housing market, this may not be the case, especially in urban areas. Research in Cardiff suggests that agency offices tend to have a more geographically defined market area, as opposed to market niche's such as flats, that is a common situation in other cities. Also, 'multi-agency' selling, where a property is on the books of more than one agency, appears to be uncommon in Cardiff. Hence, a random sample of agencies may result in sampling bias with respect to geographic coverage. Moreover, although redlining was not explicitly evident in Cardiff, anecdotal survey evidence suggested that certain agencies were selective over which areas and property types they were prepared to deal with. This reinforces the danger of relying solely on a random sample of agencies.

The unofficial data available from estate agencies and the financial institutions are published for potential house buyers in the form of individual housing sheets, or as abridged advertisements in local property newspapers. Newspapers are also the only source of information on property sales where an estate agency was not involved. Two criteria determined the type of estate agent data source used. Firstly, properties sampled had to be georeferenced to a high resolution. This meant obtaining their full address. Secondly, detailed housing attributes, including room size, had to be obtained. This information is only available from estate agents housing sheets. The other alternative, to use the readily available information published in property newspapers, was rejected since these rarely contained addresses above neighbourhood level and secondly, the housing attribute information published was variable, and none published room size. However, these newspapers were sampled to ascertain the percentage of properties sold through private transactions, without the use of an agency. The number was found to negligible, and this was confirmed by individual estate agents who put the figure at less than 5%.

Table 4.4
Estate Agent Sample Survey

Id-Number	Estate Agency	Branch	Sample	% Total
1	Abraham , Glenn & Co	Crwys Road	10	0.7
2	Black Horse Agencies	LLandaff	47	3.2
3	Charltons	Crwys Road	9	0.61
4	Chris Day & Partners	Pntcanna Road	81	5.4
7	Cornerstones	Llanishen	14	13.76
8	Cornerstones	Cowbridge Road	45	-
9	Cornerstones	Whitchurch	75	-
10	Cornerstones	Thornhill	70	-
11	Croft Davies & Co	Llanishen	6	0.4
14	Crown & Co	Cathedral Road	36	7.3
15	Crown & Co	Cowbridge Road East	11	-
16	Crown & Co	Crwys Road	9	-
17	Crown & Co	Whitchurch	6	-
18	Crown & Co	Rumney	23	-
19	Crown & Co	Llanishern	6	-
20	Crown & Co	Roath	17	-
26	General Accident	Albany Road	19	4.2
27	General Accident	Rhiwbina	6	-
28	General Accident	Llanishen	37	-
29	Geoff Edmunds	Albany Road	5	3.0
30	Geoff Edmunds	Fairwater	39	-
32	Halifax Property Services	Cowbridge Road East	47	14.3
33	Halifax Property Services	Llandaff	13	-
34	Halifax Property Services	Rumney	18	-
35	Halifax Property Services	Whitchurch	14	-
36	Halifax Property Services	Cyncoed	5	-
37	Halifax Property Services	Roath	115	-
38	Hern & Crabtree	Roath	43	2.9
39	Homeline Finance Centres	Grangetown	6	0.4
42	James Jones & Henton	Cyncoed	17	1.1
43	Jones, Michael & Co	Cathays	6	0.4
44	Keith & Co	Crwys Road	5	0.3
45	Kelvin Francis	Cyncoed	16	1.1
47	Knapp WCW & Son	Butetown	1	0
48	Knights Estate Agents	Albany Road	11	0.7
49	Lloyd Williams	Rhiwbina	2	0.1
52	Michael Graham Young	Butetown	41	2.8
55	Olney John & Partners	Cowbrigde Road East	18	1.2
56	Peter Alan	Cowbridge Road East	30	9.7
57	Peter Alan	Albany Road	41	-
59	Peter Alan	Victoria Park	22	-
60	Peter Alan	Whitchurch	22	-
61	Peter Alan	Rumney	23	-
62	Peter Alan	Rhiwbina	6	-
63	Peter Mulcahy	Albany Road	119	16.6
64	Peter Mulcahy	Cowbridge Road	106	-
65	Peter Mulcahy	Rumney	21	-
66	R.H. Seel & Co	Riverside	16	1.1
68	Rees, Barbara	Crwys Road	3	0.2
73	Sovereign	Llandaff	39	2.6
74	Taylor Simpson	Roath	28	1.9
75	Taylors	Roath	9	0.6
78	Traynor Michael & Co	Whitchurch	37	2.5
79	Wood, Thomas H	Gabalfa	11	0.7
Total			**1482**	**100.0**

4.5.2 The Estate Agent Sampling Frame.

A list of all estate agents and organisations operating within Cardiff was drawn up using *BT Yellow Pages*. This list was then edited to remove those agencies that only dealt with commercial or rental properties, and those that no longer existed. Table 4.4 is a summary of the agencies sampled, together with their location, and the number of properties sampled in the survey. The large number and range of agencies included in the survey recognises that different organisations may serve specific areas of the city, and/or different clientele. Hence the inclusion of the smaller, local independent estate agencies as well as the larger tied ones, and the building societies, banks and insurance agencies. Table 4.5 reveals that four fifths of properties were sampled from ten estate agents, some of them local independent agencies, such as Peter Mulcahy, as well as a national building society (Halifax), a bank (Black Horse Agencies) and an insurance company (General Accident). Half of the sample was surveyed from the four largest estate agencies in Cardiff, that have branches across the city.

Table 4.5

The Top Ten Estate Agents Surveyed.

	Estate Agency	Sample Size	% Total	Cumulative %
1	Peter Mulcahy	246	16.60	16.60
2	Halifax	212	14.30	30.90
3	Cornerstone	204	13.80	44.66
4	Peter Alan	144	9.70	54.36
5	Crown & Co	108	7.30	61.66
6	Chris John & Partners	80	5.40	67.06
7	General Accident	62	4.20	71.26
8	Black Horse Agencies	47	3.20	74.46
9	Geoff Edmunds	44	3.00	77.46
10	Hern & Crabtree	43	2.90	80.36

4.5.3 The Dynamic Sampling Method

The aim of the house price survey was to collect information on a cross-section of all properties in Cardiff, and to ensure that this cross-section was representative of all types of properties in all neighbourhoods. However, since properties on the market at any particular time are not necessarily representative of all properties in Cardiff, an approach had to be devised that would take this into account. Firstly, a profile of the housing stock was constructed from the 1991 Census - see Table 4.6 - and this was used as a basis for a sampling frame. Secondly, since the number of properties for sale at the time of the survey was not known *a priori,* a dynamic sampling approach was used. This recognises that the population is unknown (i.e. the total number of houses on the market), but attempts to minimise any bias caused by under - or over-sampling by building up a profile of this population based upon data already collected, and comparing this to the profile of total dwellings constructed from the Census. Thus, as the survey progressed, the type and location of properties already sampled had to be continuously monitored, and the survey modified accordingly to reflect the profile. This resulted in the concentration upon certain neighbourhoods and certain property types, such as bungalows in Cyncoed, in the latter parts of the survey since these had been under-sampled.

Table 4.6
Property Sampling Frame

Neighbourhood	Total	%	Sample	%
Adamsdown	3478	2.94	59	3.99
Butetown	1772	1.50	25	1.69
Caerau	4053	3.43	53	3.59
Canton	5789	4.90	103	6.97
Cathays	5021	4.25	58	3.92
Cyncoed	4308	3.65	44	2.98
Ely	6084	5.15	56	3.79
Fairwater	5547	4.69	60	4.06
Gabalfa	2346	1.99	20	1.35
Grangetown	5516	4.67	83	5.62
Heath	4601	3.89	36	2.44
Lisvane & St.Mellons	2349	1.99	67	4.53
Llandaff	3684	3.12	73	4.94
Llandaff North	3375	2.86	23	1.56
Llanishen	5838	4.94	102	6.90
Llanrumney	4902	4.15	36	2.44
Pentwyn	5997	5.08	67	4.53
Plasnewydd	7094	6.00	85	5.75
Radyr & St. Fagans	1897	1.61	9	0.61
Rhiwbina	4991	4.22	43	2.91
Riverside	5587	4.73	101	6.83
Roath	4384	3.71	52	3.52
Rumney	3489	2.95	22	1.49
Splott	4517	3.82	58	3.92
Trowbridge	5079	4.30	70	4.74
Whitchurch & Tongwynlais	6452	5.46	73	4.94
Total	**118150**	**100**	**1478**	**100**

Properties in the Inner Area were surveyed in a rather different manner. A street - by -street survey was undertaken, noting down properties that were up for sale, and the agencies dealing with the transaction. This type of intensive field work was performed for two reasons. Firstly, since property types and house prices varied significantly over short distances in the Inner Area, it was necessary to have a comprehensive a sample as possible. Secondly, re-assessment of street quality in the Inner Area was needed to update the one undertaken during the CHCS five years before. This re-assessment is described in more detail in the next section. It can be seen in Table 4.6 that there is a good correspondence between the percentage of total properties in each ward and the percentage of properties sampled. Wards that have been slightly over-sampled include Lisvane & St. Mellons, Llanishen and Riverside, whilst those under-sampled include Ely and Heath. This is not too unexpected, since the sample is based upon properties up for sale as opposed to total number of properties in each ward, and thus mis-matches are to be anticipated.

The dynamic sampling technique was made possible by continuously entering the address and property type of sampled properties into an Excel spreadsheet as the survey progressed. The address of the property was stored as four separate geographical fields; number, street, neighbourhood and postcode. The spreadsheet enabled the properties to be sorted by any one of these fields. Hence, the number of properties sampled in any neighbourhood or street could be determined at any time during the survey. Since a property's address could be listed in alphabetical order, it was very easy to check if a property had been previously sampled, and omit it if it had. This ameliorated the potential problems resulting from double counting, such as if the same property was on another agencies books at a different price.

Unfortunately, it was very rare for estate agents to have postcoded the properties on their books, and so these had to be added later from the Postcode Address File. Postcodes were useful for two reasons. Firstly, different estate agents appeared to have different areal boundaries for defining the neighbourhoods in Cardiff, although as discussed in section 4.2, this is a problem with the definition of neighbourhood in general. However, it was apparent that the boundaries of prestigious neighbourhoods, such as Roath and Llandaff, tended to be exaggerated by some agencies, whilst the boundaries of poorer neighbourhoods contracted. This could cause problems if the same property, or nearby properties are deemed to be in different neighbourhoods by different estate agents. Secondly, postcodes have a related grid-reference, and so the property may be plotted. The grid reference may also be used to determine the which neighbourhood the property is located, as opposed to using the estate agents more subjective description. This is discussed in more detail in the next chapter.

Using this technique, around 1500 properties were sampled during February 1995. According to estate agent sources, the market at this particular period was described as 'stable'. To gain an appreciation of the potential differences between asking and selling price, a comparison of asking price and selling price of all Cardiff properties in the Nationwide database was undertaken for 1994 and 1995. This indicated that for the majority of properties in both years (c. 90%), there was no difference between the asking and selling price at the mortgage approval stage. For the 10% of properties in 1994 that there were differences, selling price greater than asking price by an average of £3600 for 4.38% of properties, whilst selling price was less than asking price by an average of £2550 for 5.29% of properties. In 1995, these figures were 3.99% and 4.07%, resulting in an average price difference of £2350 and £2500 respectively. For the actual time period that the house price survey was undertaken (the first quarter of 1995), the corresponding figures were 3.35% and 4.83% resulting in an average price difference of £2150 and £2740. Therefore, it can be seen that for the majority of properties, there is very little or no difference between asking and selling price, and any difference is liable to be less than 5% of the eventual cost of the property. However, these differences correspond to sales financed by the Nationwide Building Society only, so the above results should be regarded within the context of the previous critique of official sources, such that the data may represent a biased sample of all housing transactions at the local level. Hence greater differences may occur for some types of properties, in some areas, although this is unavoidable.

4.5.4 The Structural Attribute Data Obtained from the Survey

Table 4.7 summaries the information on housing attributes recorded from the property sheets. In an attempt to be consistent with the data collected during the CHCS, the property attributes recorded, such as the categories used for property type and date of construction, were taken from the 1991 English House Condition Survey, as were techniques for dating types of property and estimating garden size. The latter was made simpler by the fact that estate agent literature usually commented on whether the garden was unusually different from the norm. The size of kitchen was included for three reasons. Firstly, it became evident that kitchen size tended to vary quite significantly. Secondly, many properties, and in particular flats and the linked properties, tended to have kitchen-diners as well or instead of

Table 4.7

Data Recorded in House Price Survey

Investment Property	Type of Heating
Property type	
	Gas
End-Terrace	Oil
Mid-Terrace	Solid Fuel
Semi-Detached	Electric
Detached	
Flat in Purpose Built Building	Age
Flat in Converted Building	
Maisonette	New
Bungalow	Post 1964
End-Link	1918-1964
Mid-Link	Pre-1918
Number of Bedrooms	Tenure
Size of Bedrooms	
Number of Recreation Rooms	Freehold
Size of Recreation Rooms	Leasehold
Number of Kitchens	
Size of Kitchens	Garden Size
Number of Bathrooms	
Number of Shower Rooms	None
Number of Garages	Less than 5m
Off Road Parking	5m-50m
Central Heating	Over 50m
Full	Needs Modernisation
Partial	Swimming Pool
None	Conservatory

separate dinning rooms. Thirdly, it was intended to use information on room size to construct an estimate of floor area. This would be improved with the addition of kitchen measurements. All room size measurements are in square feet. Although bathroom size was also given in some cases, this was not consistently so, and so was not recorded. The off-road parking variable represented any form of designated car parking space that was not covered by the garage category. Hence, this included driveways, carports and specified parking bays. The presence of central heating was also recorded. This was either stated explicitly in the estate agents literature, or implied by references to radiators, or the use of the boiler for central heating purposes. When neither of these were mentioned, it was assumed that the property did not have central heating installed. Two further pieces of information recorded was state of repair and whether the property was an investment property. The former represents those properties that the estate agency felt needed substantial improving, whilst the latter represented properties that were being sold with future private renting in mind, usually to a private landlord or renting agency.

4.5.5 The Street Quality Re-assessment Survey

It has been previously explained that the CHCS will provide the basis of data pertaining to locational attributes such as street quality. However, this source of data has three main problems:

- The CHCS experienced some problems regarding low response rates in specific areas. These areas were particularly concentrated in Butetown and Llandaff. Hence, there are gaps in the coverage pertaining to locational attribute information within certain localities of the Inner Area.
- Since 1989, when the CHCS was completed, several small housing developments have been built within the Inner Area, and hence no data will be available from the CHCS.
- The principle of the CHCS was to assess housing condition and environmental quality to aid future improvements. These improvements may have been undertaken since the completion of the CHCS. Also, in certain cases, street quality may have deteriorated.

Hence, a re-assessment of street quality was needed to ameliorate these problems. This was undertaken in conjunction with the Inner Area house price survey. For matters of consistency, the survey assessed street quality using the same questions as in the CHCS. The English House Condition Survey 1991 was also used during the re-assessment, since this had provided the background for the environmental quality assessments in the CHCS, and hence was a useful guide.

The quality of the local environment in the English House Condition Survey was assessed on two main dimensions: the general quality of the neighbourhood, which was termed the overall impression, and the nature and severity of specific, immediate externalities (pp. 89). The overall impression was measured in the English House Condition Survey on a seven point scale, from good (1-2), to average (3-5) through to poor (6-7), based upon the surveyor's impression of the area (e.g. Photograph 9.5. pp. 89). The main factor that appeared to influence this result was the predominant land-use of the area. At the more immediate, street level, The English House Condition Survey looked at sixteen different externalities that were regarded as 'problems'. These were eventually grouped into six main types, which are summarised in Table 4.8 and were assessed on whether the problem was of major or minor significance.

Table 4.8

Classification of Environmental Problems.

Unkempt Area	Problems with litter, rubbish, dumping, scruffy gardens, vandalism or graffiti.
Maintenance	Poor condition of roads, paths, pavements or street furniture. These problems are generally beyond the control of residents. Apart from acts of vandalism, they are not the result of the actions of residents or others.
Industry	Problems created by industry in or close to the local environment - industrial waste, pollution or noise.
Abandonment	Problems of vacant sites, non-conforming use or condition of shops / businesses.
Traffic	Problems with heavy traffic or nuisance from street parking. In some cases these problems are beyond the control of residents - through traffic or traffic caused by nearby industry and commence. In other cases they are generated by the residents themselves.
Noise	Problems with railway or aircraft noise.

Source: English House Condition Survey. 1991 Table 9.1 pp. 90. Photograph 9.6. pp. 91

The survey concluded that certain types of problems, such as graffiti, dumping and vacant sites, will have a clear impact on the local environment, whereas others, like noise from trains or aircraft or fumes from traffic or industry, may have little influence.

The criteria used in the re-assessment are summarised in Table 4.9, and were selected on the findings of the CHCS as representing those attributes that were the most important in influencing urban environmental quality in Cardiff. The condition of roads and lanes and the quality of the street environment were graded on scale from poor, to average, to above average. The remainder were graded upon whether the Environmental Health Officer who had conducted the CHCS considered that it was a problem.

As is explained in detail in *Chapter Five*, these data were aggregated to street level from individual responses in the CHCS, and were used as a yard stick against which the re-assessment took place. Before the re-assessment commenced, a sample of streets were visited so the street quality assessments in the CHCS could be gauged. These streets were chosen with prior knowledge upon whether improvements had taken place since the completion of the CHCS. On the whole, very little difference was discovered between the results recorded during the CHCS and the re-assessment. The main changes had occurred in those street that had since been renovated, and several streets in Grangetown that had deteriorated since the CHCS.

Table 4.9

The CHCS Variables Used in the Street Quality Re-Assessments

Street Quality Attribute	**Code 1**	**Code 2**	**Code 3**
Front Garden Length	Under 1 metre	1 - 5 metres	Over 5 metres
Condition of Roads	Poor	Average	Above Average
Condition of Lanes	Poor	Average	Above Average
Quality of Street Environment	Poor	Average	Above Average
Standard of Upkeep	Worse then Neighbourhood	Same as Neighbourhood	Better than Neighbourhood
	Code 0	**Code 1**	
Traffic Problem	No	Yes	
Visually Obtrusive Non-residential Landuse	No	Yes	
Atmospheric Obtrusive Non-residential Landuse	No	Yes	
Noise Obtrusive Non-residential Landuse	No	Yes	
Derelict Land	No	Yes	

Section 4.6 Conclusions

The aim of this chapter was to place the research into context by formally identify the aims and objectives, and describing the Cardiff housing market. Once this had been achieved, the remainder of the chapter investigated potential sources of data pertaining to the physical and social characteristics of the housing market. The research identified several large and complex social-economic data sets, namely the Council Tax register, the Rates register and the Cardiff Housing Condition Survey. These datasets were then complemented by a house price survey in which structural attributes of the dwelling and the asking price were recorded, and a street quality survey undertaken. The Nationwide Database was also acquired for Cardiff, covering mortgage approvals within the same time period as the house price survey. This allowed a comparison of asking and selling price to be made for similar properties in Cardiff. Therefore, a whole range of large and complex data sources have been collated, and a context-sensitive means of handling this information is thus needed. The next chapter discusses how this was achieved by the construction of two Geographic Information Systems: one for the whole of the Cardiff housing market and one for the Inner Area.

Chapter Five

Constructing a Context-Sensitive GIS

"... some degree of mismatch should be anticipated when ... independently complied data sets are brought together....[this process] is less simple than it seems at first sight" Raper, et al. (1992). pp. 86

Section 5.1 Introduction

The previous chapter has described the large property based data sets that are available for this research. The chapter concluded that there was a need for a 'context sensitive' means of handling this information. It was argued in *Chapter Three* that a GIS is an ideal medium for handling such context-sensitive information, although very little work has been done within hedonic house price research. Therefore, this chapter will describe the methodology behind constructing a context-sensitive GIS, that will take into account the hierarchical nature of the data as was suggested in *Chapter's Two & Three.* The second section shall discuss the spatial resolutions of the GIS in relation to the types of data and the purpose of the research. The next section shall then describe in detail the mechanisms used to linking the data sets to the GIS, with specific reference to the property data sets used in previous research in the Inner Area. Section four describes how this GIS was used to manipulate the data to generate urban new geographies such that the data were held at the appropriate spatial resolutions, whilst section five describes how non-residential landuses were constructed using the results of a simple modelling exercise. The final section concludes the chapter and discusses some of the implications of the GIS in the subsequent analysis.

Section 5.2 The Structure of the Cardiff ARC / INFO GIS

5.2.1 Introduction

Two separate GIS's were developed for the Cardiff study area. The first was used to research the spatial dynamics of the whole housing market, and was based upon three spatial levels of resolution; the property, the Enumeration District (ED) and the community. A second, more complex GIS, was developed for the local Inner Area study with the aim of more accurately measuring the effects of locational externalities. This GIS was based around four spatial levels of resolution that were seen as intrinsically capturing the externalities that operated at this local scale. This is described in more detail in section 5.3. This second GIS also builds upon previous research of the changing geographies of revenue raising in the Inner Area (Martin et al., 1992; Longley et al., 1993), and incorporates the property valuation data sets discussed in *Chapter Four.* Since the two GIS's were designed for different types of research and incorporated different data sets, they shall be discussed separately.

5.2.2 The Cardiff Housing Market GIS

This GIS is comprised of three spatial levels, each represented by an ARC / INFO coverage. At the individual property level, properties in the house price survey were geographically referenced via their postcode to a 100m resolution grid-reference. These grid-references were then used to generate a point coverage in ARC / INFO, which formed the basis of the GIS. The ED and community levels were added to the GIS by digitising their boundaries, generating two ARC / INFO polygon coverages. The nature of these coverages meant that all three perfectly nested - see Figure 5.1 .The level of resolution that the data are referenced and the use of census geographies to capture submarket effects, means that the GIS is very basic and essentially coarse. However, it is suitable for the analysis of the housing market as a whole. However, to model small scale locational externality effects, the Inner Area GIS has to be much more sophisticated with respect to the level of resolution that the data are referenced, and the delimitation of areal units. Hence the majority of this section is dedicated to how this GIS was constructed.

Figure 5.1

The Cardiff Housing Market GIS - The Three Nested Levels

Property Level

ED Level

Community Level

1 km

5.2.3 The Inner Area GIS

Similar to the Cardiff housing market GIS, the Inner Area GIS is constructed from ARC / INFO coverages, representing four spatial levels of resolution: the property level, the street level, the HCS area level and the community level. As was described in *Chapter Four*, HCS areas were designated by Cardiff City Council as areas within communities which had very similar housing stock and household characteristics. Given the heterogeneous nature of the Inner Area, these form a convenient additional layer between streets and communities. These areal units also reflect the small scale differences of the Inner Area's built environment, more so than communities. The latter were added as the forth level to be consistent with the Cardiff housing market GIS, and also because these represent the eleven neighbourhoods used by estate agents to divide up the Inner Area. The generation of each of these coverages are described in detail below, with particular reference to the property level coverage.

5.2.3.1 The Property Level Coverage

Previously, the Cardiff housing market GIS utilised postcoded grid-references as a means of generating a property level point coverage. However, as is discussed below, this is a very coarse method of geo-referencing data, particularly at small scales. Since the Inner Area GIS needed to be as accurate as possible to capture local level locational externalities, a different method was used to geo-reference the property level data. This method involved using the Ordnance Survey ADDRESS-POINT product as the basis of generating the property point coverage. The differences between these geo-referencing systems, and the advantages of ADDRESS-POINT shall now be discussed.

I. Postcoded Grid References

Until fairly recently, the only method for geographically referencing address-based data to a reasonable resolution was by the use of its postcode. All postcodes in the UK are stored in the Central Postcode Directory (CPD), and the Postcode Address File (PAF), created by the Post Office (Raper et al, 1992). Both data sets contain a record of every unit postcode in the country (c 1.7 million) and a corresponding national grid reference, although the PAF also contains a definitive listing of the UK's 25 million postal addresses. The bulk of the national grid references are actually given to sufficient digits to provide a 10m ground resolution, but in practice, the last digits are zeroes and hence the real resolution is 100m. The convention adopted was that the co-ordinates corresponded to the 100m south west corner of the OS national grid containing the first address in the postcode. Traditionally, there has been concern about the accuracy of these grid-references since the CPD was not created with detailed geo-referencing in mind (Gatrell, 1989; Raper et al 1992). Research by Gatrell and others have demonstrated inaccuracies in the CPD relating to grid-references, and have highlighted inherent problems caused by the method of assigning co-ordinates. For instance, in urban areas many different postcodes share the same grid references, whilst in rural areas a number of properties dispersed across a wide area may share the same postcode and hence grid reference. It is also quite common for odd and even numbered properties in the same street to have different postcodes, and the convention of assigning grid references has often resulted in these properties having different 100m co-ordinates. In addition, since the first address in the postcode is used as the source of reference, it is not unusual for the bulk of the properties in a unit postcode to lie in the neighbouring 100m grid. A general conclusion from such research is that the grid references are far from optimal and in some cases are a cause for concern (Raper et al. , 1992). Until recently, the only alternative to the postcoded grid reference was the Pinpoint Address Code (PAC), a high resolution product developed in the late 1980s. PAC was more accurate than the CPD since it assigned a 1m resolution grid reference to each individual property. However, for various reasons, this data set failed to become a national property-level referencing system (Martin, 1992).

II. Ordnance Survey ADDRESS-POINT

ADDRESS-POINT was launched by the OS in 1993, and similar to PAC, provides a National Grid co-ordinate for each address, but instead to a resolution of 0.1m (Ordnance Survey, 1993). Unlike PAC, ADDRESS-POINT has also achieved national coverage, and is constantly updated. The product was created by matching the PAF against Ordnance Survey Land-Line digital database, using Ordnance Survey Centre Alignment of Roads (OSCAR) information as a guide. In most cases the property seed point from the Land-Line is then assigned as the address location, with a resolution of 0.1m (ibid.). A unique Ordnance Survey ADDRESS-POINT Reference (OSAPR) code is also given for each separate postal address in the PAF. In some cases, separate postal addresses in the PAF will share the same seed point; for instance, different addresses in the same building. This means that two dwellings on different floors of the same building may share a common grid reference, but will have unique OSAPRs. However, where there is only one delivery point in the building, and mail is sorted internally, there will be only one OSAPR relating to one property seed. The OSAPR applies even when an address is changed, or the original structure is demolished and replaced. It only becomes 'dead' if the structure is demolished and not rebuilt.

In the first version of ADDRESS-POINT, used in this research, a few unresolved addresses were assigned a lower resolution grid co-ordinate, such as those recorded in the CPD. This occurred if there was no apparent match between the PAF and Land-Line, and manual matching by surveyors in the field was unsuccessful. A status flag indicates the resolution of the seed point, as well as its type, physical status and match status of the address reference. Seed type specifies whether the structure is a permanent or temporary building; physical status indicates whether the building is planned, existing or demolished; and match status indicates the degree of agreement between the PAF and the Ordnance Survey databases concerning the specification of address. Hence each ADDRESS-POINT record has four components: the full postal address, a grid-reference, an OSAPR, and a quality statement. For the first time, ADDRESS-POINT represents a high resolution, fully comprehensive national property-level referencing system for the UK (Martin et al, 1994).

5.2.3.2 The Street Level Coverage

The street level coverage represents the second tier of the Inner Area GIS. The line coverage was generated by digitising the whole street network for the Inner Area of Cardiff (Martin et al, 1992). The coverage contains approximately 920 individual streets, and each was given a unique identification code based upon the street codes extracted from the unique property reference numbers (UPRNs) in the register (see section 5.3 for more detail). The street names were also attached to the corresponding arcs in the coverage. For the purposes of calculating accessibility measures, this line coverage was enlarged with the addition of main roads that connect the Inner Area to the M4 at the periphery of Cardiff.

5.2.3.3 The HCS Area and Community Level Coverages

Two polygon coverages were generated, representing the eighty one HCS and eleven communities that comprise the Inner Area. Each polygon in the community coverage was assigned its ward code taken from the 1991 census and also its unique identification code extracted from the UPRNs in the rates register (see section 5.3 for more detail). Each HCS polygon was assigned a HCS identification code, taken from the CHCS.

Section 5.3 Linking Spatial Information to the GIS

Chapter Four has described the various address based data sets that are used in the research. This section will discuss the mechanisms and procedures utilised to incorporate these data sources into both the GIS for the whole Cardiff housing market and the Inner Area GIS. Since the GIS for the whole housing market is less sophisticated, and will contain less data, this will be described first. The remainder of the section though will describe the methods for attaching data to the Inner Area GIS

5.3.1 Linking Data to the Cardiff Housing Market GIS

Two data sources were incorporated into the GIS covering the whole Cardiff housing market: the house price survey data and the 1991 census data. The former were attached to the property level point coverage generated from the postcoded grid references. This was achieved by using the JOINITEM command with the postcode as the relational join since this was common to both the point coverage and the house price survey. This allowed the full set of structural attributes to be attached to the GIS. However, in some cases, more than one property shared the same grid-coordinate for reasons explained in section 5.2.

The 1991 Census data at the ED level were attached to the ED polygon coverage using the ED census code that was common to both. Two methods were then considered in linking these data to the property data in the point coverage. The first involved using the POINT-IN-POLYGON (PIP) analysis techniques to overlay the ED polygon coverage on top of the property point coverage, and linking the data based upon which properties fell into which ED. However, this was not used because of the relatively low resolution and accuracy of the grid-referenced data which may have placed a large number of properties in the wrong ED, especially in urban areas where mean ED size is small (Boyle et al, 1991; Martin, 1991). Instead the ED to postcode directory was used to match the census data to the postcode coverage. The ED to postcode directory was developed by OPCS so that postal geography could be incorporated into the 1991 Census. By means of the directory, the postcodes associated with any of the 112 000 EDs in England and Wales may be ascertained. On average, an ED will be associated with 14 unit postcodes (Martin, 1996). Since many unit postcodes will fall across ED boundaries, Psuedo-EDs (PEDs) where constructed. A PED represents the ED in which the majority of the current (1991) population living within the unit postcode falls. Hence PEDs were used as these are the nearest approximation to an overlap between unit postcodes and EDs.

Therefore the two data sets were incorporated into the Cardiff GIS. The community level coverage remained devoid of attribute data. However, each property in the point coverage was assigned one of twenty six community identification codes, which were used in the subsequent analysis.

5.3.2 Linking Data to the Inner Area GIS

It has already been discussed that in the Inner Area GIS, different attribute data will be held at different spatial resolutions. Table 5.1 describes this in more detail. This indicates that the property level coverage will hold data pertaining to property valuations and structural attributes, as well as locational attributes relating to accessibility and proximity to non-residential landuses. The street level coverage will contain street environment data, and also data concerned with secondary school catchment areas. The HCS area level coverage will hold data relating to the local environment, particularly the quality of local amenities, whilst the community level coverage will contain data pertaining to social composition.

Table 5.1. A Summary of Housing Attributes in each Coverage

Property Level Attributes	Property valuation data Rateable value Council tax band House price Structural attributes Accessibility measures to work place Proximity measures to non-residential landuses.
Street Level Attributes	Street environment measures Class of street Street quality Non-residential activity School catchment areas
HCS Area Level Attributes	Housing density Quality of local amenities
Community Level Attributes	Social composition

Figure 5.2
Examples of Address Formats in some of the Different Datasets

An Extract from the Pre-1990 Rates Register

UPRN	RATEABLE VALUE	ADDRESS
10101200024102	67	HOUSE 24 MARY STREET CARDIFF
10102000021208	161	MAISONETTE 21 JESSICA STREET CARDIFF
10102000039108	161	MAISONETTE FIRST & SECOND FLOORS 39/40 JESSICA STREET CARDIFF
10109000003056	67	FLAT(THIRD FLOOR) 3/4 TYLER STREET CARDIFF
1010900002640A	57	FLAT (FOURTH FLOOR) 26 TYLER STREET CARDIFF
10109000028072	60	FLAT (FIFTH FLOOR) 28 TYLER STREET CARDIFF
10109000028053	82	FLAT (THIRD FLOOR) 28 TYLER STREET CARDIFF

An Extract from the Council Tax Register

UPRN	COUNCIL TAX BAND	ADDRESS
101000012000240	C	24, MARY STREET, CARDIFF, CF1 5AB
101000020000210	B	FLAT 1ST FLR AT 21, JESSICA STREET, CARDIFF, CF1 1FG
101000020000391	B	1ST & 2ND FLR AT 39/40, JESSICA STREET, CARDIFF, CF1 1FF
101000090000030	A	3/4, TYLER STREET, CARDIFF, CF1 2AW
101000090000260	A	26, TYLER STREET, CARDIFF, CF1 2AW
101000090000280	A	28, TYLER STREET, CARDIFF, CF1 2AW
101000090000281	A	28, TYLER STREET, CARDIFF, CF1 2AW

An Extract from Ordnance Survey ADDRESS-POINT

OSAPR	ADDRESS
AP16M18L1BD45QL011	24, MARY STREET, CARDIFF, CF1 5AB
APTAY94BH4467PL011	FLAT 1, 21, JESSICA STREET, CARDIFF, CF1 1FG
APKLM178BC457PN011	39A, JESSICA STREET, CARDIFF, CF1 1FF
APQ16NL3BDD45QL011	3, TY-LER STREET, CARDIFF, CF1 2AW
APTAY94SH4467PL011	26, TY-LER STREET, CARDIFF, CF1 2AW
APQ1N18LBCD45QL010	FLAT 1, 28A, TY-LER STREET, CARDIFF, CF1 2AW
AP16N18L3BD45QL011	FLAT 2, 28B, TY-LER STREET, CARDIFF, CF1 2AW

A pre-requisite to this matching exercise was undertaken in previous research discussed in Martin et al, (1992); and Longley et al, (1993). This research investigated methods of matching diverse urban data sets such as the rates register, electoral roll, council tax register, and the CHCS and attaching this to the digitised street coverage of Cardiff. Since this forms the foundations of the matching techniques described in the next section, a brief summary of the methodology from this work shall now be discussed.

5.3.2.1 Matching the Address Based Datasets.

Most Local Authorities hold property level data sets that, although are not usually explicitly geo-referenced, contain addresses in some coded format (Higgs et al, 1995). This is true for addresses in the rates register, the council tax register and the CHCS, which all have a UPRN. Frequently, a UPRN may be in a coded format corresponding to sub-divisions of a property, its dwelling number, its street address, and its community location. These so called 'intelligent' UPRNs are useful since a property's location may be systematically approximated from these codes. This is an advantage since many such data sets do not contain postcoded addresses. This situation applies to the rates register and the CHCS, although not to the council tax register which is fully postcoded.

Two techniques were considered when matching the three registers. The first involved matching using the address of the property. However, matching text strings is particularly problematic given the variety of different address formats and conventions in existence in administrative data sets. Figure 5.2 illustrates this by providing examples of the address formats found in the rates register, council tax register and the ADDRESS-POINT data set. This situation is improving with the launch of the British Standard BS7666 which specifies a standard format for address and property referencing, although it will be some years before administrative data sets reflect this (Cushnie, 1994). Since a variety of address formats and street and property naming standards were found to exist in the different registers, address based matching proved difficult.

The second technique involved matching by using the UPRN, and is summarised by a three stage process in Figure 5.3. In the first stage, the UPRNs in each register had to be standardised, since they varied in their formats. This was possible since they all had the same common ancestry (the rates register), and so could be deciphered and converted into a

Figure 5.3
The Procedure for Matching the Three Registers

Corresponding UPRNs in the rates register, council tax register and CHCS

Record Number	Rates Register	Council Tax Register	CHCS
1	10101200024102	101000012000240	101120000240
2	10102000021208	101000020000210	*
3	10102000039108	101000020000391	*
4	10109000003056	101000090000030	101900000030
Total number of records	46159	45028	7413

STAGE ONE
Standardisation of UPRNs Between Data Sets

Record Number	Rates Register	Council Tax Register	CHCS
1	101000012000240	101000012000240	101000012000240
2	101000020000210	101000020000210	*
3	101000020000391	101000020000391	*
4	101000090000030	101000090000030	101000090000030

STAGE TWO
Linking Data Sets by UPRNs

	Rates Register	Council Tax Register	CHCS
Total number of matched records	38462	44001	6477

STAGE THREE
Decomposition of UPRNs into Properties, Streets and Communities

Record Number	House Number	Street-ID	Community-ID	Rateable Value	Council Tax Band	CHCS Data
1	24	120	101	67	C	....
2	21	200	101	161	B	*
3	39	200	101	161	B	*
4	3	900	101	67	A	....

common format using a custom written FORTRAN programme. This allowed the addresses in each register to be matched in a second stage using another FORTRAN programme. However, some discrepancies occurred during this matching process, due to differences in the content of each register, particularly with respect to sub-divided properties. This is illustrated in Figure 5.2, with respect to the dwellings in **'Jessica Street'**, in which the sub-divided properties are coded differently in each register. As such, the coding for some sub-divided dwellings were unresolvable, and mismatches occurred. This resulted in loss of data, which is shown in stage two of Figure 5.3. The other main reason for mismatches was the existence of an address in only one of the registers. This was particularly true for the rates register, which contained addresses that had become obsolete over time. The UPRNs were also 'intelligent' in as far as it was possible to break them down into separate codes that related to the community, street and dwelling number of the property. These codes allowed each property to be grouped together into communities and streets in the final stage using a third FORTRAN programme. Hence, a file was created that contained the identification code of the property, together with a street and community identification code, the rateable value, the council tax band and where appropriate, information from the CHCS.

5.3.2.2 The Original Street Based GIS

A 'loosely-coupled', street based ARC / INFO GIS (Martin et al, 1992) was then constructed by attaching this matched file to the digitised street coverage of the Inner Area, and the polygon ward coverage of the eleven communities in the Inner Area. This was possible since the community and street codes extracted from the UPRNs had been incorporated into each coverage as they had been digitised. This allowed the matched file to be joined to the GIS using the JOINITEM command with the codes as the relational join. Hence, this GIS was capable of handling address based information down to the level of the individual street. This GIS was then subsequently used to investigate the changing geographies of local revenue raising in the Inner Area. However, in terms of being locationally sensitive, the GIS had several drawbacks, particularly with respect to the different levels of resolution. For instance, property level data could only be manipulated at the resolution of the street, seriously influencing the resulting analysis. Also, since the street coverage did not nest into the HCS or community coverages, data had to be interpolated from streets that crossed these boundaries. These types of problems are indicative of the geographical anarchy that has grown over the years, in the sense that most British data are assembled for sets of areas, but these areas are generally inconsistent (Raper et al., 1992). Hence, the next section will describe how the above GIS was improved by the addition of a property level coverage, made possible by the integration of ADDRESS-POINT data. This proved to be the major part of constructing the Inner Area GIS.

5.3.3 Linking Property Information to the Inner Area GIS

ADDRESS-POINT information was obtained from the Ordnance Survey for the twenty postcode sectors covering Inner Cardiff, amounting to 63634 properties. These were subsequently matched with the three data sets following exclusion of those ADDRESS-POINT properties outside the Inner Area.

5.3.3.1. ADDRESS-POINT and the Council Tax Register

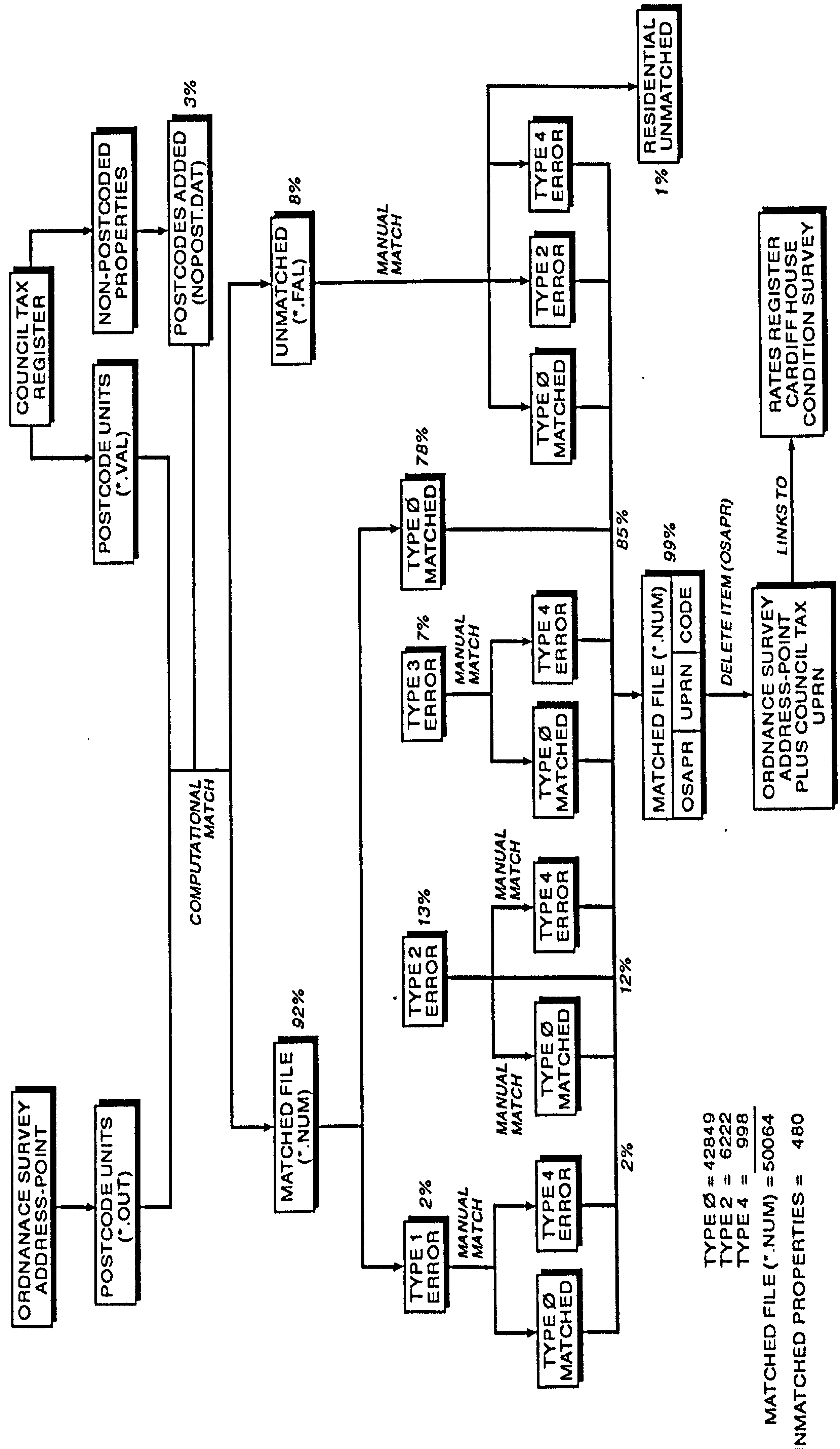
"[O]ne particular way of enhancing the quality of data is to make comparisons between two independently collected versions ... this needs to be investigated" Raper et al, (1992) pp 84

The aim of the matching was to provide a mechanism whereby the address in the February 1994 version of Ordnance Survey's ADDRESS-POINT could be related to the matched database consisting of the council tax register, the rates register and the CHCS. Unfortunately, ADDRESS-POINTs version of a UPRN, its OSAPR, is not 'intelligent' and hence cannot be broken down into constituent spatial codes for streets or communities. Hence a different matching mechanism had to be used. After investigating the address formats of each data set and for ADDRESS-POINT, two procedures were considered:

1. the matching of ADDRESS-POINT and the data sets by text-strings;
2. the matching of ADDRESS-POINT and the data sets by a mixture of postcodes and addresses

The issues surrounding the first technique have previously been discussed (section 5.3.1). Although the address format within ADDRESS-POINT correspond to the new British Standard, this is not the case for the other registers, and hence the same problems still apply. The latter technique offers the possibility of isolating unit postcodes from both ADDRESS-POINT and one of the registers and, for each unit postcode, using the number of the

Figure 5.4
A Summary of Matching the Council Tax Register with ADDRESS-POINT



dwelling as the 'match' between the two data sets since each unit postcode contains approximately 12 - 14 addresses (Raper et al, 1992). This will allow properties that have differing text based information, such as '**16, Briony Street, CF2 5AD**' and '**Flat 16 Briony St. CF2 5AD**', to be matched. Since the only data set to be fully postcoded was the council tax register, this data set was used in the matching procedure. A FORTRAN programme was written that isolated the dwelling number and unit postcode of the addresses contained in both the council tax register and ADDRESS-POINT, along with the corresponding UPRN and OSAPR, and these were saved in two separate files. Of the total number of properties on the council tax register for the Inner Area in April 1991 (45658), just over 1.3% (c 650) were not postcoded. Postcodes were subsequently added to the files using the PAF. A second FORTRAN programme then matched the UPRNs to the OSAPRs using the postcode and dwelling number as the matching criteria.

5.3.3.2 Results of the Matching Procedure

"[T]he Royal Mail's task is certainly complicated by the fact that many properties ... have names, not numbers, and are amalgamations of former individual properties" Raper et al, (1992). pp. 80.

The result of the matching of the two data sets are presented in Figure 5.4, which illustrates the complexity of the matching process as well as highlighting the types of errors likely to result from such matching. As a result of matching addresses at the individual unit postcode level, two files were produced. One represents situations where a match between the two files were found. The other contains addresses for which no match was found. This situation arose when:

1. An address existed in the council tax register, but not in ADDRESS-POINT.
2. An address in the council tax register had the wrong postcode (either unit, sector or district);
3. An address in the council tax register consisted of the name and not the number of the property or similarly, where the property in ADDRESS-POINT was so referenced;
4. An address in the council tax register had a composite number such as 213-215 or 213/215 making a straight forward match on a single numeric problematic, especially where the ranges were seen to vary.

5. Instances when the form of the address in the council tax register could not be related to the layout of the address in ADDRESS-POINT (e.g. **'Front Gnd Flr Flat 1 at No. 73 Gabriel Street'**) or where addresses contain alphanumeric descriptions (e.g. 47A). This situation arose where the same flat, for example, is described in different ways in the two registers (e.g. basement flats being described as lower ground floor flats, third floor flats being described as Top flats, etc.).

The first three types of matching problems are caused by the form of the addresses in the council tax and ADDRESS-POINT data sets. The latter two are the results of the programming procedures. In all, a match was found for just over 92% of the properties in the council tax register. However, the matched file also contained addresses for which matches were found but where there were problems in the matching procedure. These were flagged in the matched file:

A **Type 0** match occurred where there was a perfect 'one-to-one' match (accounts for just over 78% of the matched file)

A **Type 1** match accounted for just under 2% of the matched file and occurred when multiple OSAPRs were matched to one UPRN. Instances of this occurred where a property had been sub-divided into constituent dwellings in the ADDRESS-POINT information (for example, **'14 and 14 A Rattigan House'**) but is present in the council tax register as **'14 Rattigan House'**. The programme will effectively match on the number 14 twice, resulting in this instance in the two dwellings being assigned the same UPRN.

A **Type 2** match accounted for just under 13% of all the matches and occurred when one OSAPR was matched to multiple UPRNs. This exists where there is a sub-divided property in the council tax register but only one delivery point for the building (i.e. a Housing of Multiple Occupation (HMO)). Hence, there would only be one ADDRESS-POINT location co-ordinate and one OSAPR. These errors also occurred with the reverse of Type 1 matches.

A **Type 3** match, which accounted for just under 7% of matches, occurred where multiple OSAPRs were matched to multiple UPRNs. These situations primarily occurred where a block of flats have the same postcode. For example, the properties in **'Flats 1-10 Chelsea Court CF2 3RT'** and **'Flats 1-10 Diana Towers CF2 3RT'** will all have unique UPRNs and unique OSAPRs. But since the number and the postcode of each property appears in the same combination twice, the programme will match two UPRNs to two OSAPRs. This situation would also arise where multiple dwellings in the council tax register were being matched with multiple entries in the ADDRESS-POINT file such as where four flats in **'Bannerman Avenue'** are matched with what appears in ADDRESS-POINT as two sub-divided properties. Instances also occurred where properties had been given the incorrect postcode in the council tax register and have therefore been matched incorrectly. This distinction is important since failures of the former are caused by the programme design and can be adjusted manually, whereas the latter are due to inconsistencies in the council tax register which are more problematic.

Type 2 matches, in general, cannot be resolved unless associated with multiple Type 3 matching errors. In such cases they were been corrected. Where Type 1 and Type 3 were resolved manually they were given the code of 0 (i.e. a perfect match). Those remaining properties were been given a code of 4. In such cases, a UPRN cannot be assigned to an individual OSAPR due to the differing sub-divisions of the property existing in the two data sets or due to variations in the property description. Manual checking of the errors led to an increase of 7% in the perfectly matched file (code = 0). The residual Type 2 and Type 4 matching errors accounted for 12% and 2% of the 'matched file' and cannot be resolved without recourse to more detailed field work.

Of the 8% of properties in the council tax register that had failed to be matched by the computer programme, only 480 of them could not be subsequently matched manually, representing just under 1% of the properties in the Inner Area council tax file. These include addresses that:

1. Exist in the council tax register but not ADDRESS-POINT

2. The address in the council tax register having a name as opposed to a number and number in ADDRESS-POINT (or vice versa) making matching impossible without recourse to detailed field investigation

As a result of the matching process, the matched file contains three status codes for each property where:

Type 0 represents instance where properties have been matched exactly

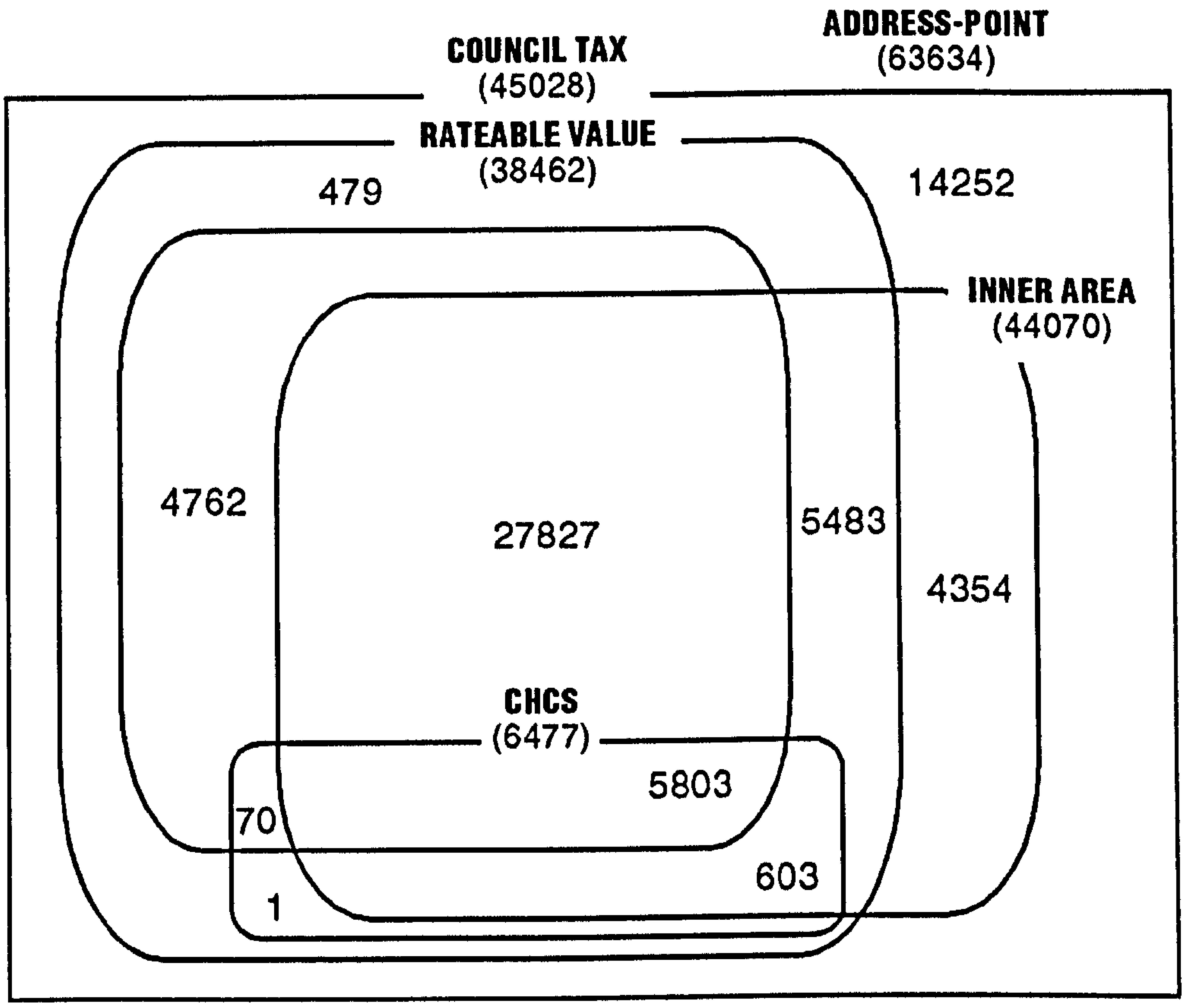
Type 2 represents cases where matches had been achieved but for HMOs such that the same OSAPR had been attached to multiple UPRNs. These were flagged in the database but cannot be investigated in further detail.

Type 4 represent cases where a match cannot be achieved on an individual address due to the inherent complexities of the differing nomenclature within the two data sets. Although these properties have been retained in the final matched file, and an OSAPR has been assigned to 'approximate' the correct reference, only detailed field investigations could reveal the 'exact' OSAPR. A Type 4 could then be used to identify the location of these properties. As such, these properties only differ from the unmatched file because in the case of the latter it has not been possible to even approximate the property location because of the non-compatibility of the two registers for certain properties.

In all, the computer matching and manual procedures were able to resolve matches for 99% of the council tax addresses, but these 'matches' incorporate a number of uncertainties. A direct 'perfect match' was achieved for 85% of addresses. In 2% of cases more than one OSAPR corresponds with a single UPRN. In these circumstances, sub-division of HMOs have been captured in the ADDRESS-POINT data, but not in the council tax register. This is not necessarily an obstacle to the geo-referencing of the existing register. The inverse situation of multiple UPRNs for the same OSAPR account for 12% of the 'matches'. Again, this allows a location to be assigned, but is illustrative of the differences of the definition of properties between the two systems. Of these various error types, manual checking resulted in considerable improvements, but was very time consuming. Mismatches are spread throughout the study area, but are particularly concentrated in those communities with a large proportion of sub-divided property such as Cathays.

Similar conclusions were discussed by Raper et al (1992), but with respect to the inaccuracies of the PAF. These included properties found on the ground that were not in the PAF, such as recently constructed houses, 'dead' properties that had not been removed from the PAF, such as demolished houses, and errors occurring with house numbering in 'complex areas'. These were often affluent neighbourhoods where improvement and fusion of properties had occurred. Since the PAF formed the basis of ADDRESS-POINT, it is hardly surprising that similar problems were discovered. Moreover, recent work undertaken

Figure 5.5
Summary of the Data Loss after the Matching Process



by the National Land Information Service (NLIS) has highlighted similar problems when matching address based data sets to ADDRESS-POINT for properties in Bristol (Smith, 1996). Thus information loss appears to be a common occurrence with data set integration, and has a tendency to increase with the more data that are added. This is illustrated in Figure 5.5, which summaries the data loss when the three registers were matched to ADDRESS-POINT. However, the value added by the integration of these data sets more than outweighed that lost through mismatches.

The resulting matched file was imported into the ADDRESS-POINT coverage of the Inner Area GIS. This was achieved by the JOINITEM command between data held within the ADDRESS-POINT property coverage and the matched file, since the OSAPR was common to both. This subsequently allowed the rates register and the CHCS to be joined to ADDRESS-POINT via their UPRN's.

5.3.3.3 Incorporating the House Price Survey Data

The house price survey data for properties in the Inner Area were linked to the ADDRESS-POINT property level coverage in a similar manner as the above data sets. However, since the number of Inner Area properties in the house price survey was much smaller, these were matched to ADDRESS-POINT manually rather than using a computer programme. This was achieved efficiently by using Excel to search for the address text strings in ADDRESS-POINT, and copying the OSAPR into the house price spreadsheet. However, similar problems to the computer matching were still encountered, specifically problems with HMOs (Type 2 error), and sub-divided property (Type 1 error). An additional problem were properties sampled in the house price survey that did not exist in ADDRESS-POINT. These were typically newly constructed properties. The house price survey data were then attached to the ADDRESS-POINT coverage using the JOINITEM command with the OSAPR as the relational join.

Section 5.4 Generating New Inner Area Urban Geographies

Once the data from the three registers and the house price survey were attached to the ADDRESS-POINT property level coverage, the Inner Area GIS coverages were then linked together so that data could be aggregated to different levels of spatial resolution. Before this was undertaken, however, the ADDRESS-POINT property level coverage was cleaned to remove all those properties that did not have a quality statement implying a resolution of 0.1m. These included those properties whose co-ordinates had been extracted from the CPD and thus were only referenced to a 100m resolution. In all, 4.7% (2068) of properties were excluded. These included 69 properties in the CHCS and 19 properties in the house price survey. An analysis of these excluded properties suggested an over representation of sub-divided properties and HMOs, which is understandable given their ambiguous nature in the PAF.

5.4.1 Linking Coverages in the Inner Area GIS

Manipulation of the data within and between each of the four spatial levels was an important concept underlying the construction of the Inner Area GIS. Hence, the property, street, HCS area and community coverages had to be linked together. This was achieved via the system of unique identification codes that were stored in each coverage. The idea was to match these unique identification codes from one coverage to another, such that a point in the property level ADDRESS-POINT coverage contained the unique identification codes for the property, street, HCS area and community that it was located within. Hence, the concept was very similar to an 'intelligent' UPRN, although in this case four individual reference codes were stored as opposed to one. This linked-GIS gives a certain flexibility when attaching data to the spatial hierarchy. For instance, each coverage may only contain level-specific data. So at the property level, the point coverage may only be related to property level housing attributes. At the street level, the line coverage may only have street level externalities, whilst at the HCS area level, the polygon coverage may only have HCS area level externalities. Alternatively, a pyramidal structure may be desired. Here the point coverage of individual properties may not only contain the structural attributes of individual properties but also all the street level and HCS area level and community level locational data. By the same procedure, the street level coverage may also contain higher level locational data. This flexibility would allow different data to be generated and different models estimated. Also, the use of point data as a basis of the GIS allows the user to have control over the subsequent higher spatial units. However, before the GIS could be linked, the coverages in each spatial level were required to nest perfectly into each other, and this was the key problem in the previously described 'loosely-coupled' street-based GIS used on the previous research.

5.4.2 Nesting the Four Spatial Levels

Nesting was feasible in the case of the ADDRESS-POINT coverage, since the individual properties were capable of nesting perfectly into both streets, HCS areas and communities. In fact, the unique HCS area identification codes had already been attached to each property in the ADDRESS-POINT coverage, since these had been recorded in the CHCS data set. In addition, the unique street identification codes were linked to the ADDRESS-POINT property level coverage by the use of the street codes that had been built into the street coverage during the construction of the street-based GIS (see previous discussion). These codes had been isolated from the rates register UPRN. Since this register had been attached to ADDRESS-POINT by the above matching procedures, it was an easy process to match the unique street identification code in the street-network coverage to each property in the ADDRESS-POINT coverage.

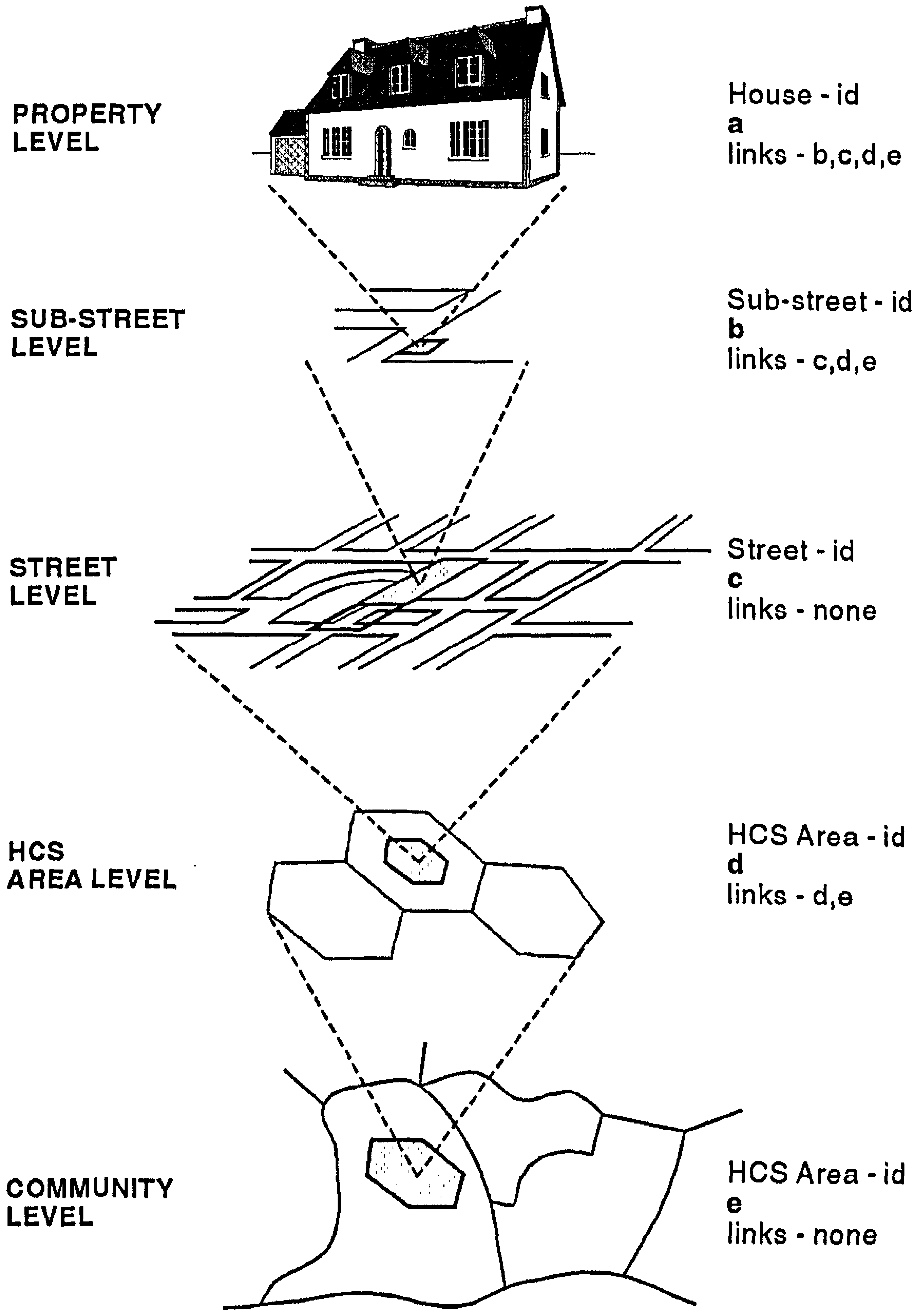
Nesting the street coverage into the HCS areas proved to be problematic, since many of the longer streets traversed the HCS area boundaries. There was also the additional problem that the street network had been used as a template for the construction of the HCS area boundaries. This had resulted in many of the HCS area boundaries running down the centre of a street, and hence the difficulty of allocating such a street into a specific polygon. The problem of nesting streets into HCS areas was overcome by the construction of 'sub-street' sections. These were the sections of street that fell wholly into each HCS area. In most cases (73%), the street and sub-streets were the same, since they both fell within one HCS area. However, in some cases (27%), a single street was split into more than one sub-street, either because it traversed a HCS area boundary or because the boundary ran down the centre of the road. In the case of the latter, properties on opposite sides of the road were in placed into different sub-streets since they were in different HCS areas. Since HCS areas nested perfectly into communities, there was no difficulty of nesting sub-streets into communities.

5.4.3 Generating Sub-streets

Sub-streets were generated using the ADDRESS-POINT coverage as a basis since all the properties had already been perfectly assigned to both streets and HCS areas, and had thus been allocated the corresponding street and HCS area identification codes. These identification codes were used to create a new set of unique identification codes indicating the sub-street section in which the property was located. This was achieved by exporting the

Figure 5.6

The Five Nested Levels of the Cardiff Inner Area GIS



property, street and HCS area identification codes to a file and using a custom written FORTRAN programme to construct the new sub-street identification codes based upon the combination of street and HCS area identification codes. These new sub-street identification codes were then attached to the ADDRESS-POINT property coverage, using the property identification code as the relate item.

Hence, each property in ADDRESS-POINT was spatially referenced by five unique identification codes: its property identification code, its sub-street identification code, its street identification code, its HCS area identification code and its community identification code. With the exception of the street, this spatially referenced system nested perfectly into four hierarchical layers. This is demonstrated in Figure 5.6. In this example, five unique identification codes (a,b,c,d,e) are stored for a property in the ADDRESS-POINT coverage, representing the property-id, sub-street-id, street-id, the HCS area-id and community-id. The arc in the sub-street network coverage stores four identification codes (b,c,d,e), whilst the HCS area coverage will store two identifiers (d,e). It should be noted that the street level has no ~~has no~~ links to any of the other levels.

5.4.4 Constructing Sub-street Level Data from the CHCS

The CHCS contained data relating to the street environment (see *Chapter Four*, section 4.5 and Table 4.9), although these were held at the property level. Using the linked GIS, these data were aggregated up to the sub-street level. One of the reasons for doing this was because the CHCS only represented 20 % of the properties in ADDRESS-POINT, and only 18 % (124) of the properties sampled during the house price survey. Hence, it was necessary to extrapolate the data from the CHCS to the remaining properties in the ADDRESS-POINT property coverage. This is justifiable in the case of street level data, since the data extracted from the properties sampled in the CHCS will also apply to all the properties in the same street. It was decided to use sub-street sections as the basis for the aggregation as opposed to actual street sections, since many of the longer streets in Cardiff will experience significant changes in street quality along their course. The modal value of each street quality variable in Table 4.9 was then calculated for each sub-street section. These were then used to create two sub-street quality attributes; one measuring overall environmental quality, the other measuring the presence of non-residential landuse in the sub-street. The first attribute was constructed by totalling together the first six street quality attributes in Table 4.9, to produce an indice of quality than ranged from six (worst) to fifteen (best). These were then divided into four categories representing general variations in street quality across the Inner Area; poor (6-7), below average (8-10), above average (11-13) and good (14-15). The second attribute was constructed by totalling the remaining four attributes in Table 4.9, to create an index that measured the presence and obtrusiveness of non-residential land use in the street. This ranged from zero (not present) through to four (very obtrusive).

5.4.5 Constructing HCS Area Level Data from the CHCS

Using similar principles to the sub-street level coverage, data pertaining to local amenities in the CHCS were extracted and aggregated to the HCS level and linked to the HCS area coverage. Table 5.2 is a summary of this data. It should be noted that due to the resolution of the aggregation, there were very little problems encountered due to missing data, since the sample of properties in the CHCS were far greater at the HCS level than at the street level, even in those areas that had new residential or Local Authority housing developments. Also, due to the nature of the attributes, there was less of a need for the data to be re-assessed and updated.

Table 5.2.

Neighbourhood Data taken from the CHCS

Neighbourhood Attribute	**Coded 1**	**Coded 2**	**Coded 3**
Quality of local shops	poor	average	good
Quality of public transport	poor	average	good
Quality of sports centre	poor	average	good
Quality of local parks	poor	average	good
Quality of community facilities	poor	average	good

5.4.6 Linking Census Data to the Community Level

The 1991 census data at ward level (as described in *Chapter Four*, section 4.4) were attached to the community coverage using the ward code that was common to both as the relational join. The ward code had been linked to the property level ADDRESS-POINT coverage by POINT-IN-POLYGON analysis techniques. This provides an accurate matching mechanism since, unlike Cardiff GIS, all the properties are geo-referenced to a 0.1m resolution, and hence will fall within the correct ED.

Section 5.5 Constructing the Geography of Inner Area Landuses

5.5.1 Introduction

The previous sections have described the construction of a four tiered context-sensitive GIS for the Inner area of Cardiff. Although the Inner Area GIS contains a wealth of data pertaining to property and locational attributes, it contains generally very little information that could be used to generate proximity measures, particularly to different landuses. Hence, the Inner Area GIS was enriched with additional coverages identifying the location of amenities that are hypothesized to generate locational externalities. There is also no reason to presuppose that externalities beyond the boundary of the Inner Area will have no effect upon house prices within the Inner Area. Hence, these additional coverages also contained externality producing landuses beyond the designated Inner Area.

5.5.2 Modelling the Locational Externality Effects

As a pre-requisite to constructing externality generating landuse coverages, the theoretical influence of location upon property prices was explored by investigating the relationship between council tax band and rateable value of each property. Previously, it has been argued that rateable value can be equated to the use value of a property whilst council tax band is more akin to the capital value. As such, it could argued that any differences between these two valuations are primarily due to the effects of location. The relationship between the council tax and rateable value was modelled using multi-nominal logit regression, with the council tax band as the dependent variable. This was achieved by converting the council tax banding system (A-H) into a numerical equivalent that took into account the ratios between the bands - see Table 5.3. The principal results of this modelling exercise was an investigation of the residuals for each property. Positive residuals occur when the model under-estimates the council tax band and negative residuals occur when the model

Figure 5.7

The Council Tax and Rateable Value Residuals

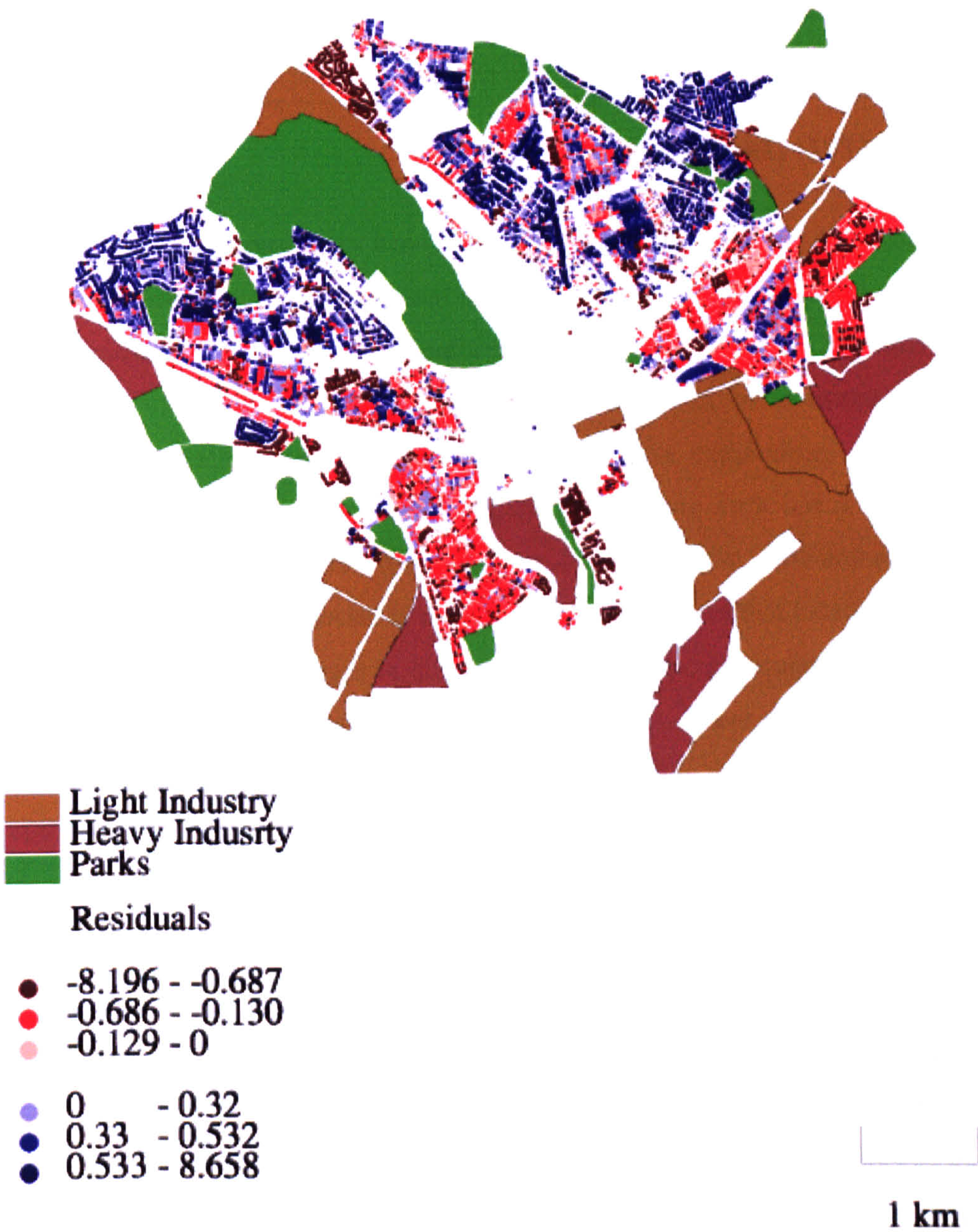


Table 5.3
Welsh Property Values, Valuation Bands and Ratios to Base Tax

Council Tax Band	House Price Bands	Ratio to base tax
A	up to £30 000	6/9
B	£30 000 to £39 000	7/9
C	£39 000 to £51 000	8/9
D	£51 000 to £66 000	9/9
E	£66 000 to £90 000	11/9
F	£90 000 to £120 000	13/9
G	£120 000 to £240 000	15/9
H	£240 000 and above	18/9

Source: Longley et al., (1993) pp. 88. Table 1.

over-estimates the council tax band. Under the assumption that the unexplained variation is caused by unaccounted for locational differences, mapping the residuals may indicate areas where positive and negative externalities exist. Figure 5.7 reveals that there is a definite geography of positive and negative residuals, which can be interpreted as positive and negative externalities respectively. Generally, positive externalities are dominant in the north and west of the Inner Area, suggesting that Bute Park has a strong influence on property prices, and negative externalities are dominant in the south and east, suggesting the possible influence of the docks. In more detail, it can be seen that localised differences seem to be associated with changes in non-residential landuse, particularly with respect to parks and open space, industrial areas and commercial areas. There also seems to be small scale linear features associated with railway lines. These results were used to inform the types and classifications of landuse coverages generated in the GIS. .

5.5.3 Constructing Landuse Coverages

Using the results from the above modelling exercise, the Inner Area's non-residential landuses were divided into three main categories: parks and open space, industrial areas and institutional landuses. In accordance with the interpretation of the residuals, the parks and open space category was further sub-divided to make the distinction between Bute Park, which appeared to have a particularly dominant influence upon property prices. In a similar

Table 5.4

Inner Area Non-residential Landuse Classifications

Park / Openspace	Area (km^2)	Light Industry (cont.)	Area (km^2)
Butetown Open Space	0.03	Colchester/Dominions Way Industrial Estate	0.21
Channel View Playing Fields	0.05	Clydesmuir Industrial Estate	0.10
Coronation Park	0.01	East Moors Trading Estate	0.82
Grange Gardens	0.01	Jubilee Trading Estate	0.04
Jubilee Recreation Ground	0.02	Portmanmoor Industrial Estate	0.41
Lansdowne Hospital Fields	0.11	Ocean Park trading Estate	1.65
Lawrenny Avenue Playing Fields	0.09	Hadfield Road Trading Estate	0.13
Mill Gardens	0.01	Leckwith Industrial Estate	0.20
Moira Place Gardens	0.01	**Heavy Industry**	
Moorland Park	0.02	Ace Industrial Estate	0.22
Ninian Park Athletics Ground	0.03	Butetown Works	0.25
Roath Park	0.11	Leckwith Industrial Estate	0.16
Sevenoaks Park	0.05	Queen Alexandra Dock	0.39
Splott Park	0.06	Seawall Road Industrial Estate	0.36
Thompson Park	0.05	**Institutional**	
Tremorfa Park	0.13	Maindy Barracks	0.17
Victoria Park	0.06	Crown Way Government Offices	0.31
Waterloo Gardens	0.01	City Hall	0.14
Waterloo Recreation Ground	0.05	H.M Prision	0.03
Light Industry		Cardiff Institute of Higher Eductaion	0.02
Gabalfa Industrial Estate	0.29	Cardiff University	0.44
Ipswich Road Industrial Estate	0.12	Welsh Office	0.27
Selan Industrial Estate	0.13	Cardiff Arms Park	0.08

manner, industrial landuses were sub-divided into 'heavy' and 'light' industrial areas. The former corresponds principally to the traditional manufacturing and extractive industries associated with the docks, since the residuals implied that these have a particularly negative influence upon property prices. The latter principally corresponds to modern trading estates that are generally devoid of these traditional types of industries. Institutional areas form the final major landuse category. These correspond to non-residential buildings and activities, such as the Government Offices and the University. Furthermore, Table 5.4 shows how these landuses have been further differentiated by their size. This will be taken into account when externality measures are calculated in *Chapter Six.*

Table 5.5

Non-residential Landuse Point Coverages

Non-residential Landuse	**Number**
Community Centres	16
Hospitals	6
Local Shopping Centres	8
Primary Schools	30
Secondary Schools	8
Sports Centres	11

Table 5.5 summarises other non-residential landuses in the Inner Area that are hypothesized to have an effect upon property prices, but weren't identifiable in the residual map. Point coverages were constructed to capture their locations. In addition, since previous research has suggested that the influence of secondary schools may also been determined by size of the school and educational attainment of the pupils, information on pupil size and examination results were acquired for all the secondary schools in the Inner Area from the publicly available School and College Performance tables - see Table 5.6. This information shall be used in *Chapter Six* when generating hypothetical school catchment areas. Finally, the location of the railway lines and the River Taff were added to the Inner Area GIS. A point coverage representing the location of all the seven railway stations that serve the Inner Area, and a separate coverage consisting solely of Cardiff Central, Cardiff's principal railway station, were also constructed. Figure 5.8 shows the resulting composite GIS, illustrating how the built environment of the Inner Area can be abstracted.

Figure 5.8

A Map of the Inner Area Built Environment

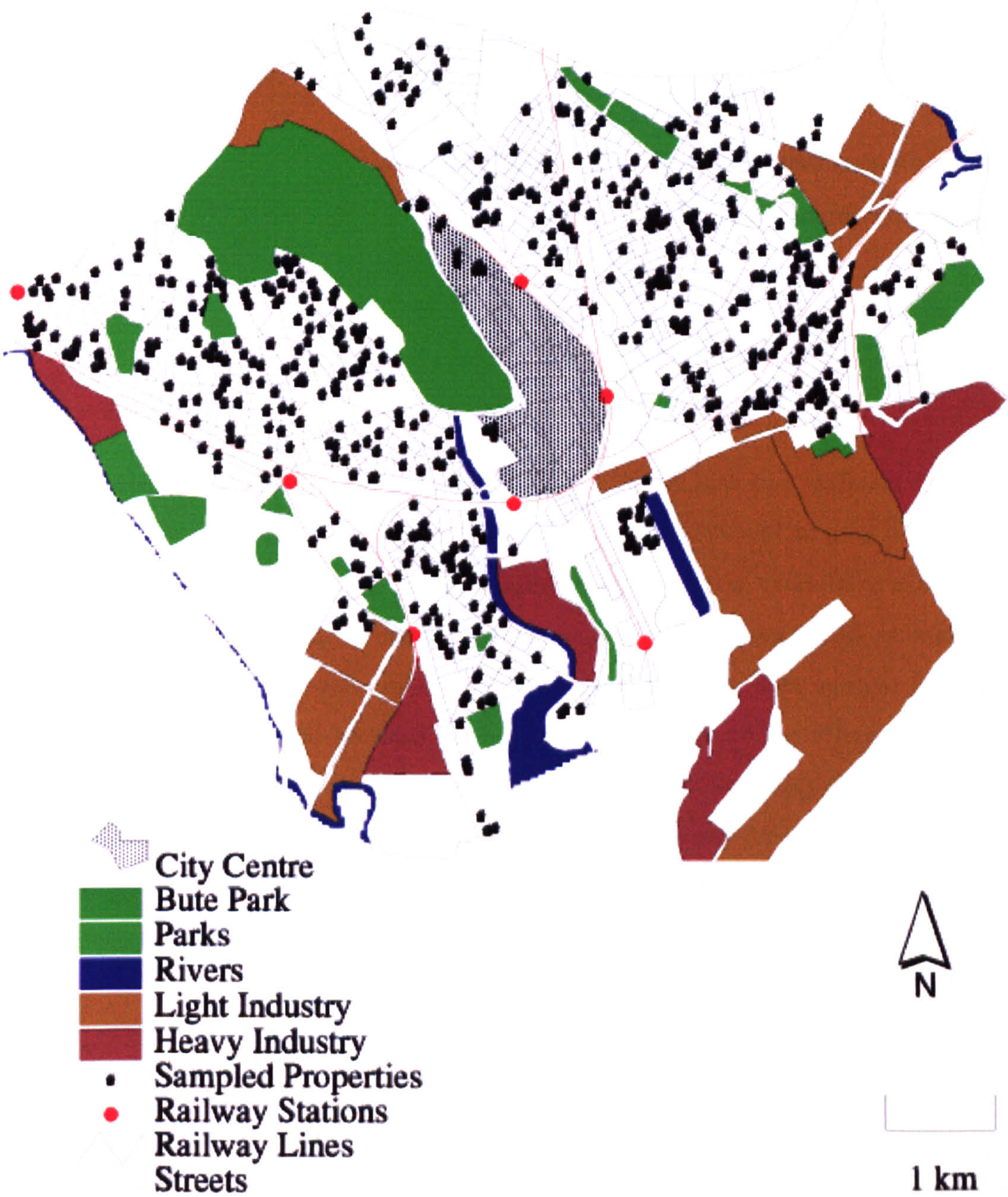


Table 5.6

Secondary School Performance Indicators

School	Type	% GCSE 5+ A-C	Total Pupils aged 11-16	% Days Absent
Cantonian High School	Local Authority	22	1003	17.9
Cathays High School	Local Authority	13	889	14.6
Fitzalan High School	Local Authority	15	1292	15.6
Howell's School,	Independent	100	403	5.5
Kings Monkton School	Independent	57	146	7.7
New College School	Independent	77	140	5.8
St Teilo's C.I.W. High School	Local Authority	49	823	8.2
Willows High School	Local Authority	23	785	20.4

Source: School and College Performance Tables 1993

Department Of Education

Section 5.6 Conclusions

To summaries, this chapter has described the methods of how two different GIS's were constructed for use in the subsequent research. The Cardiff GIS will be used in the analysis of the spatial dynamics of the whole housing market, and is essentially a basic point coverage containing all the property related data. The Inner Area GIS is much more sophisticated, and is capable of manipulating and analysis data across four spatial resolutions. This GIS will be used in the detailed analysis of the effects of locational externalities. Hence, with the GIS's complete, the research is now ready to utilise the GIS to explore the data and generate the locational attributes necessary for the subsequent hedonic modelling. This is undertaken in *Chapter Six*, which uses the various tool boxes in ARC / INFO to generate new spatial data, particularly with respect to the Inner Area.

Chapter Six

GIS and Exploratory Data Analysis

Section 6.1 Introduction

The aim of this chapter is three-fold. Firstly, it will describe and explore the structural attribute data, with particular emphasis upon the inter-relationships between property size and the other attributes. The second section will examine how locational attribute data were generated using the GIS for both the Cardiff housing market and the Inner Area study. The final section concerns the estimation of preliminary hedonic models in an investigatory capacity, and describes a more in-depth exploration of the housing data. The chapter concludes with the formulation of a standard hedonic model that will be developed in the subsequent chapters.

Section 6.2 The Structural Variables

6.2.1 Description of the Variables

Table 6.1 summarises the structural variables constructed from the structural attribute data held in the property level coverage (see Table 4.7 in *Chapter Four*). For future reference, the abbreviations of all the housing attributes, both structural and locational, can be found in *Appendix One*. Two measures of floor area were calculated from the room size information. Total floor area was estimated by summing together the size of each bedroom, recreation room and kitchen in a property. Although this measure will be less than the true floor area, it is arguable that it is the size of the habitable rooms that is significant in determining house price. The second measure relates to the average floor area of bedrooms, recreation rooms and kitchens in a property. The remaining variables directly relate to the structural attributes in the property level coverage.

Table 6.1

The Structural Variables

Variable	Key	Description
Total Floor Area (sq-ft)	Floor Area	Continuous
Average Bedroom Floor Area (sq-ft)	Ave Bed	Continuous
Average Recreation Room Floor Area (sq-ft)	Ave Rec	Continuous
Average Kitchen Floor Area (sq-ft)	Ave Kit	Continuous
Dwelling Type		
End-Terraced	ET	Dummy
Mid-Terraced	MT	Dummy
Semi-Detached	SD	Dummy
Detached	D	Dummy
Flats in Converted Building	FCB	Dummy
Purpose Built Flats	FPB	Dummy
Maisonette	M	Dummy
Bungalow	B	Dummy
End-Link	EL	Dummy
Mid-Link	ML	Dummy
Number of Bedrooms	Beds	Count
Number of Recreation rooms	Recs	Count
Number of Bathrooms	Baths	Count
Number of Shower rooms	Showers	Count
Full Central Heating	Full CH	Dummy
Partial Central Heating	Part CH	Dummy
Gas Central Heating	Gas	Dummy
Number of Garages	Garages	Count
Off-Road Parking	ORP	Dummy
Age: New	New	Dummy
Age: Post 1964	Post 1964	Dummy
Age: 1918 - 1964	1918-64	Dummy
Age: Pre-1918	Pre-1918	Dummy
Garden: None	Gdn: None	Dummy
Garden: Less than 5 metres	Gdn: < 5m	Dummy
Garden: 5 - 50 metres	Gdn: 5-50m	Dummy
Garden: More than 50 metres	Gdn: > 50m	Dummy
In need of modernisation	Needs Mods	Dummy
Swimming Pool	Swm Pool	Dummy
Conservatory	Con	Dummy

Figure 6.1
Distribution of Floor Area

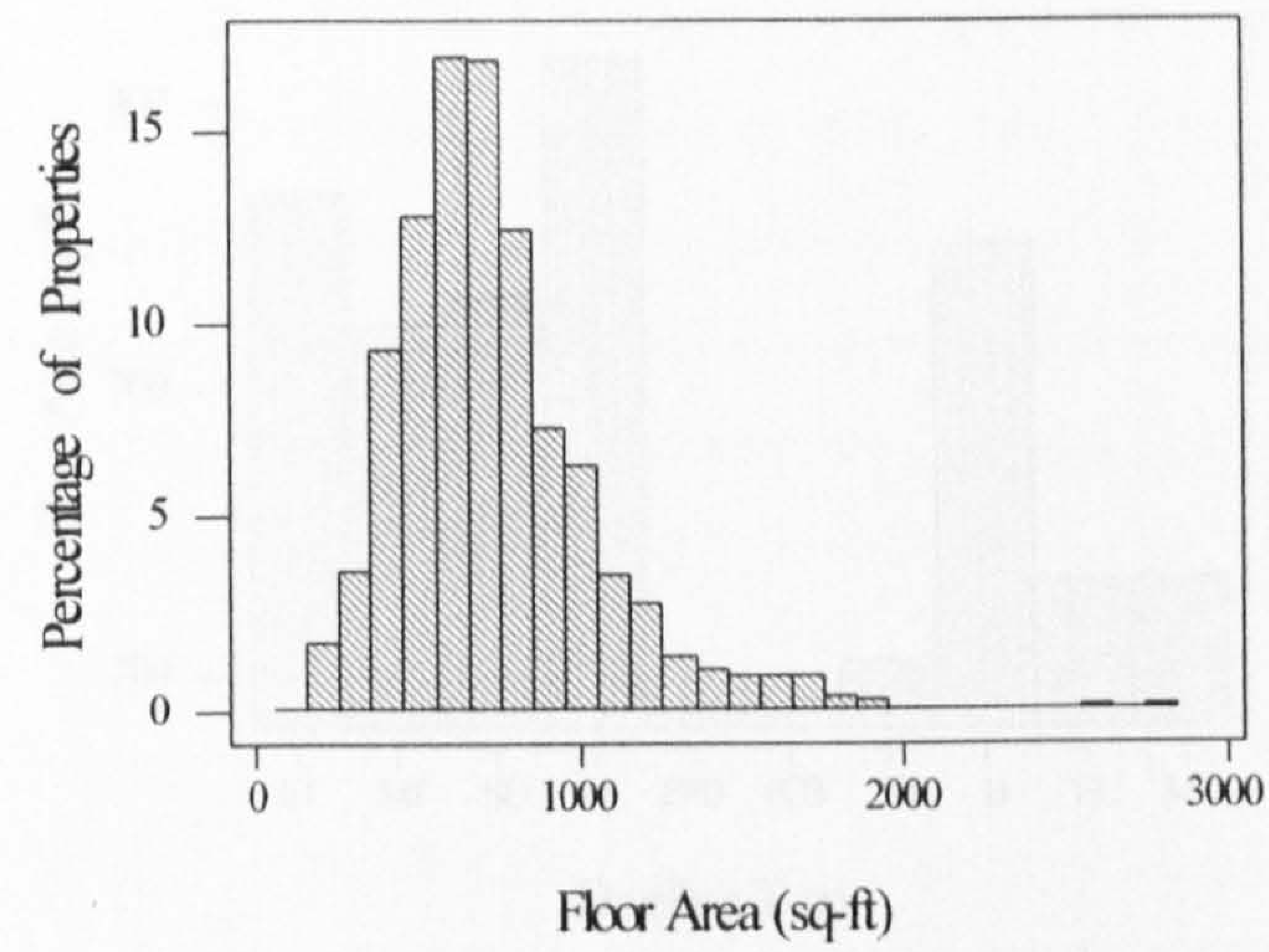


Figure 6.2
Mean Total Floor Area by Dwelling Type

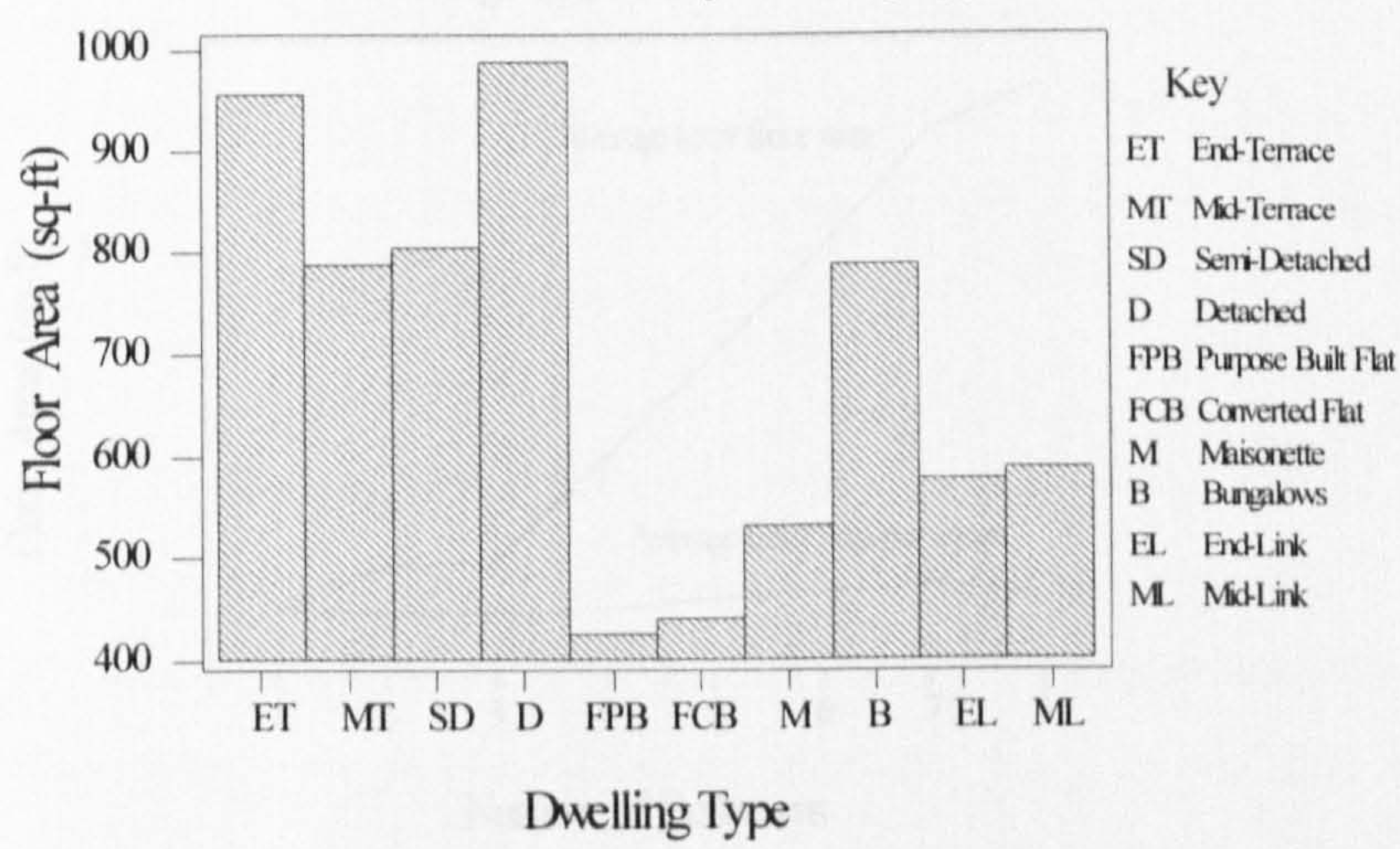


Figure 6.3
Mean Bedroom Floor Area by Dwelling Type

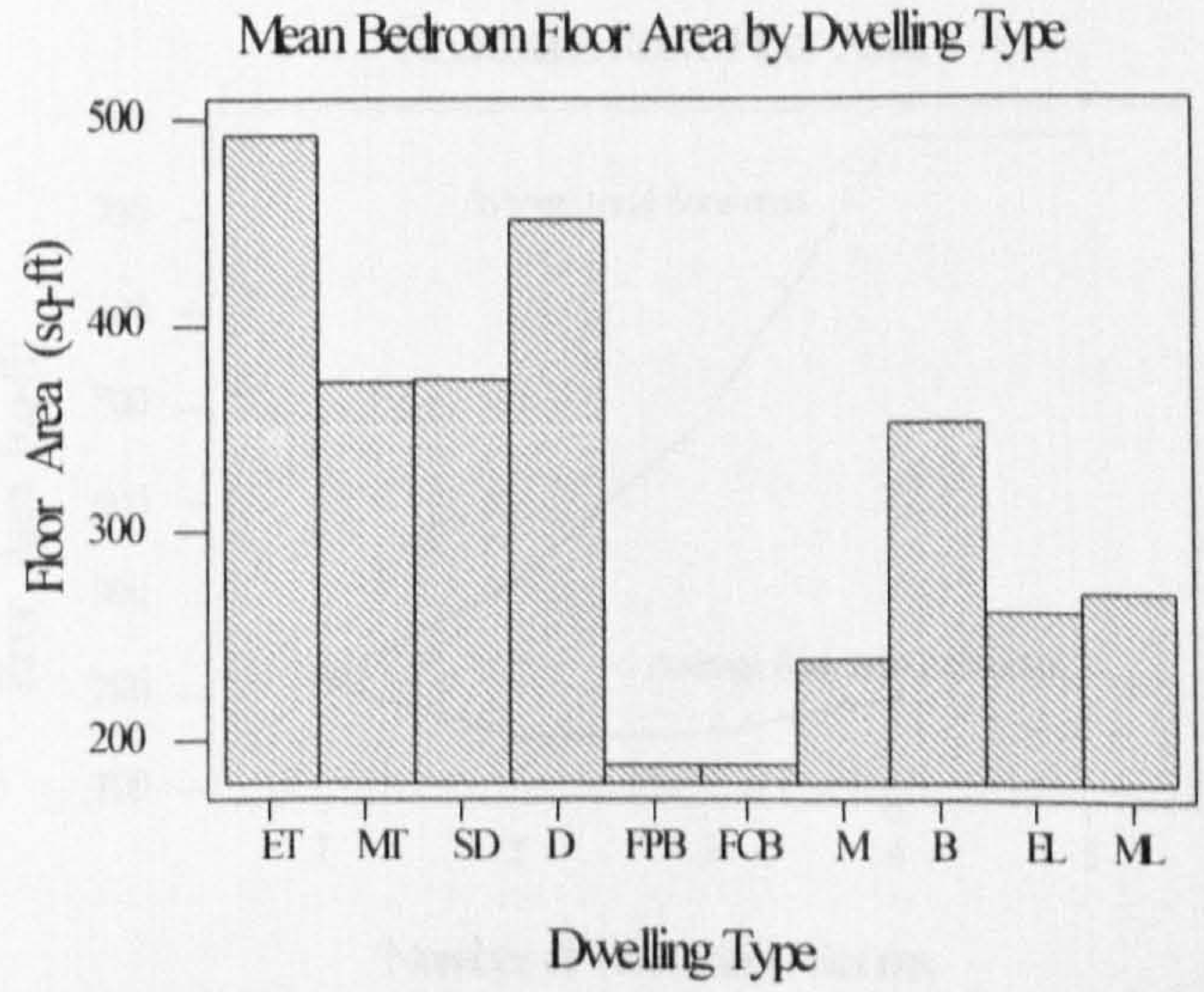


Figure 6.4

Mean Recreation Room Floor Area by Dwelling Type

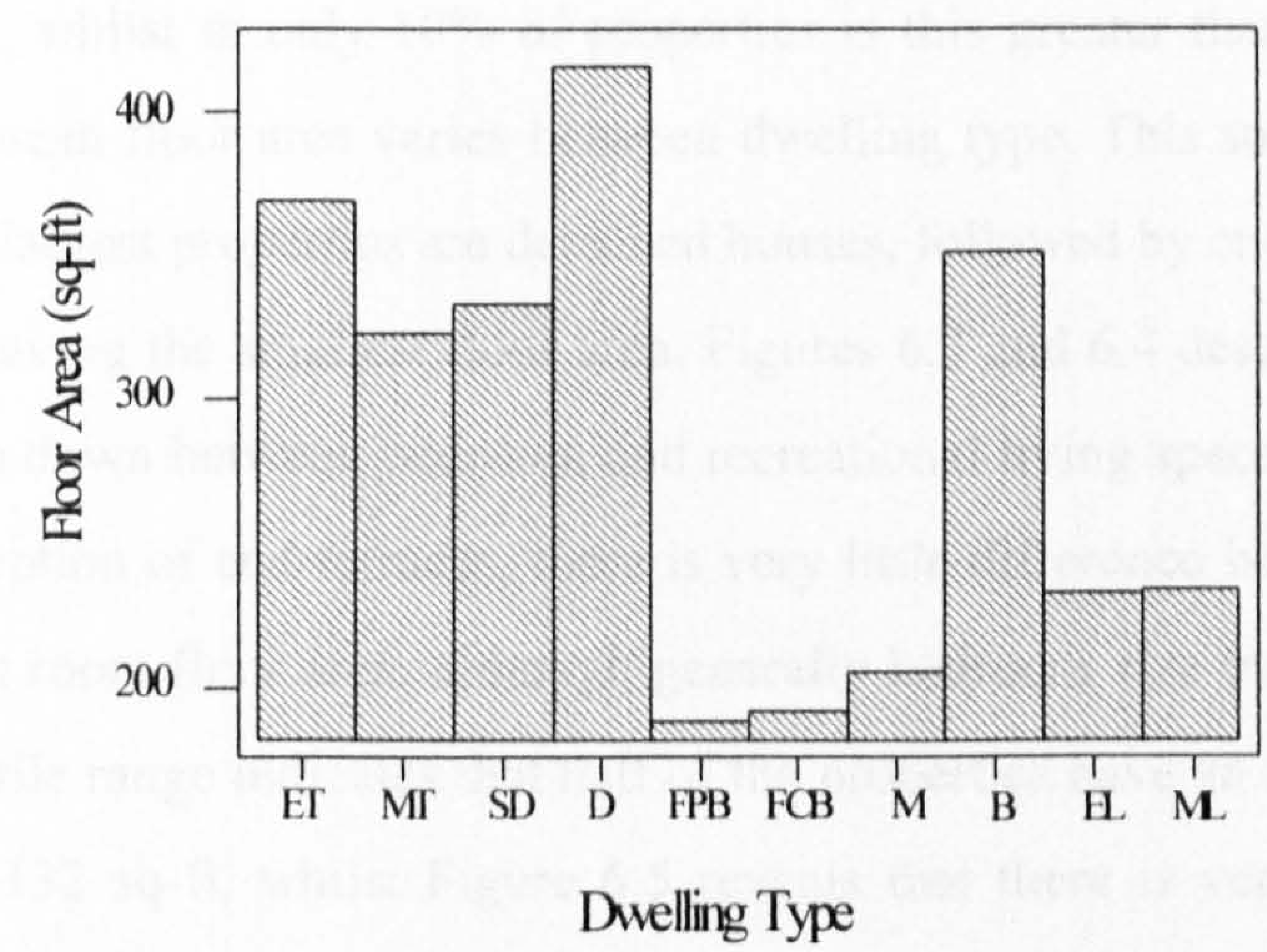


Figure 6.5

Average Bedroom Floor Area

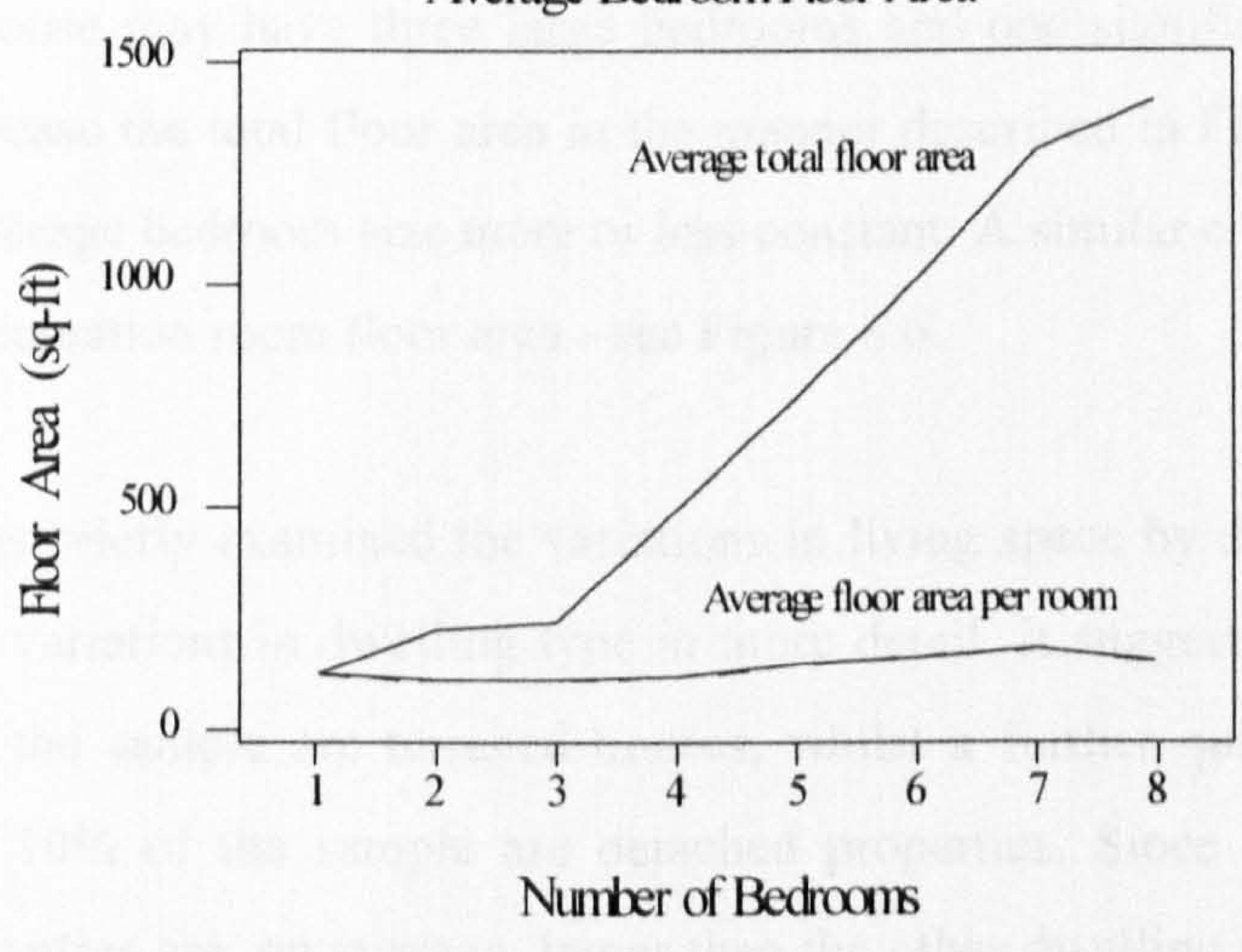


Figure 6.6

Recreation Room Floor Area

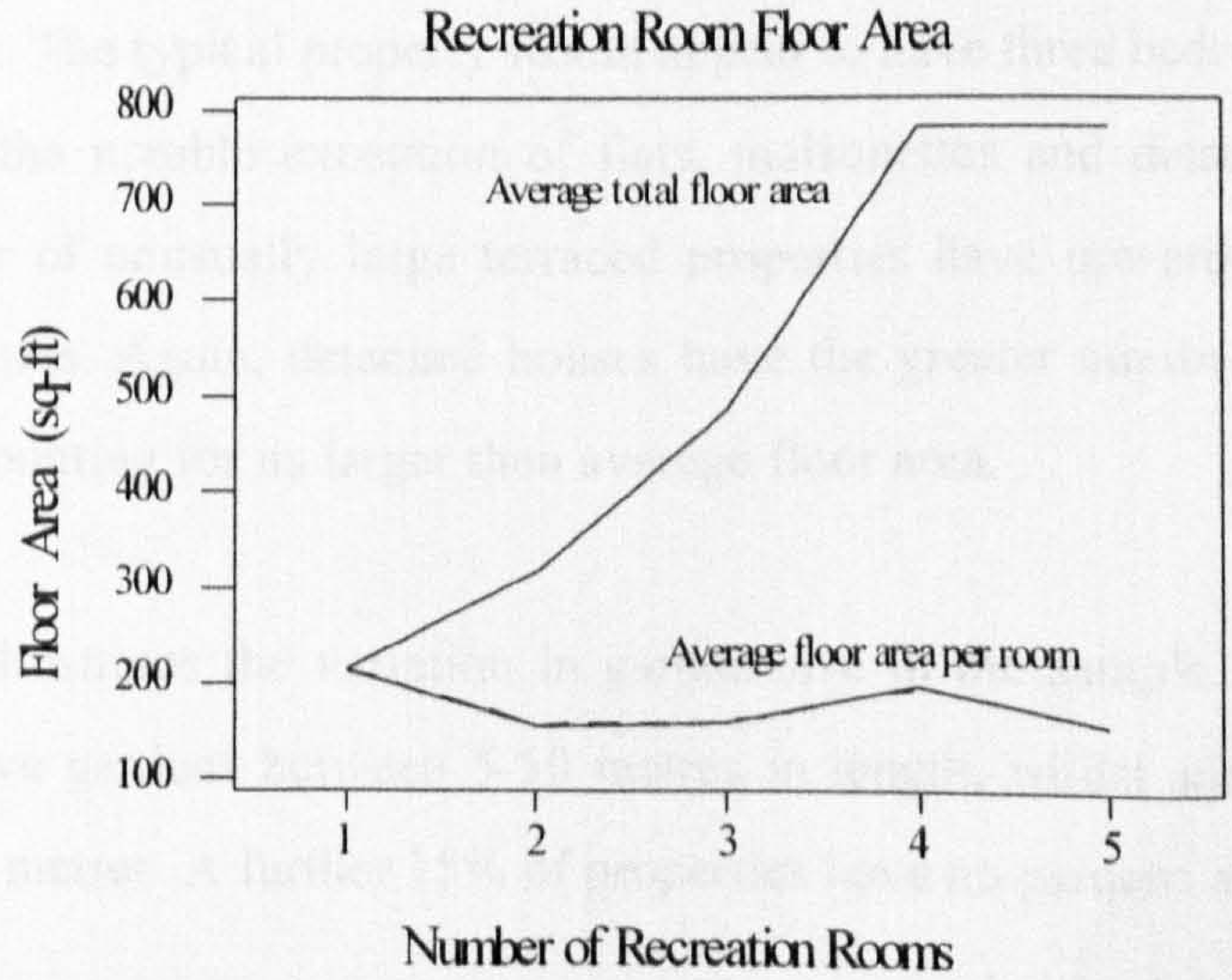


Figure 6.1 shows that total floor area follows an approximate normal distribution, and is slightly negatively skewed. Nearly half of all the properties have a floor area of between 550-850 sq-ft, whilst in only 10% of properties is this greater than 1000 sq-ft. Figure 6.2 reveals how mean floor area varies between dwelling type. This suggests that, as would be expected, the largest properties are detached houses, followed by end-terraces, with flats and maisonettes having the smallest floor area. Figures 6.3 and 6.4 describe how this total floor area is broken down between bedroom and recreational living space for each dwelling type. With the exception of end-terraces, there is very little difference between average bedroom and recreation room floor area, although generally bedroom size tends to be slightly larger. The interquartile range indicates that half of the properties have an average bedroom size of between 101-132 sq-ft, whilst Figure 6.5 reveals that there is very little variation in this average as the number of bedrooms per property increases. However, the same graph shows that total bedroom floor area rapidly increases in properties with four or more bedrooms, suggesting that bedroom size in larger houses is more variable. For instance, a four bedroomed house may have three large bedrooms and one significantly smaller bedroom. This will increase the total floor area in the manner described in Figure 6.5, whilst keeping the overall average bedroom size more or less constant. A similar conclusion can be reached with regard recreation room floor area - see Figure 6.6.

The above has briefly examined the variations in living space by dwelling type. Figure 6.7 describes the variations in dwelling type in more detail. It suggests that almost of third of properties in the sample are terraced houses, whilst a further quarter are semi-detached. Only around 10% of the sample are detached properties. Since Figure 6.2 implied that detached properties are, on average, larger than the other dwelling types, it can be assumed that these constitute the majority of the properties in the positively skewed tail of Figure 6.1. Figures 6.8 and 6.9 describe the distribution of bedrooms and recreation rooms in each dwelling type. The typical property would appear to have three bedrooms and two recreation rooms, with the notable exception of flats, maisonettes and detached properties. A very small number of unusually large terraced properties have upwards of five bedrooms and recreation rooms. Again, detached houses have the greater number of rooms for a typical property, accounting for its larger than average floor area.

Figure 6.10 illustrates the variation in garden size in the sample. Just under a half of all properties have gardens between 5-50 metres in length, whilst another third have gardens less than five metres. A further 15% of properties have no gardens at all. Figure 6.11 reveals

Figure 6.7
Variations in Dwelling Type

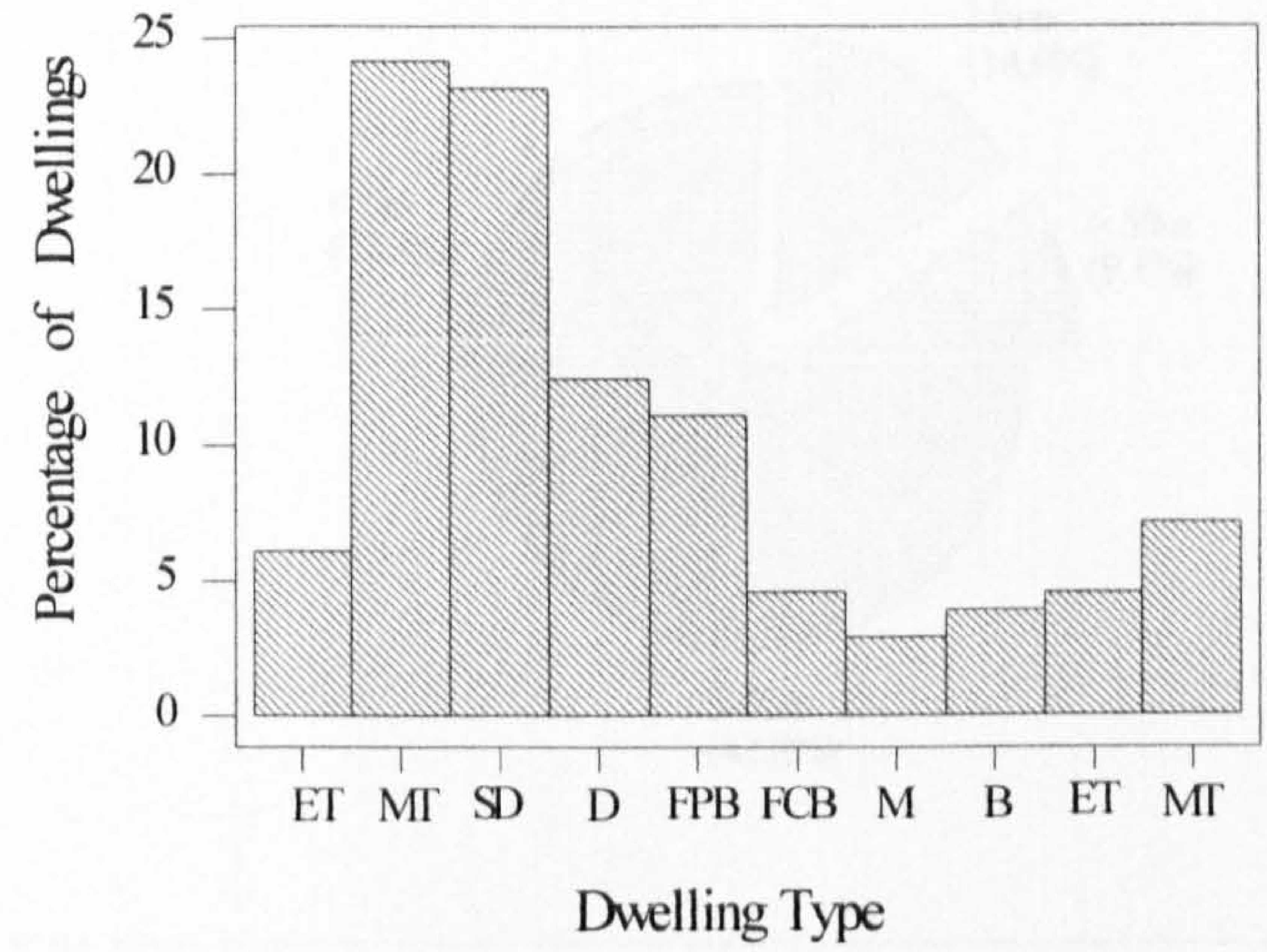


Figure 6.8
Number of Bedrooms in each Dwelling Type

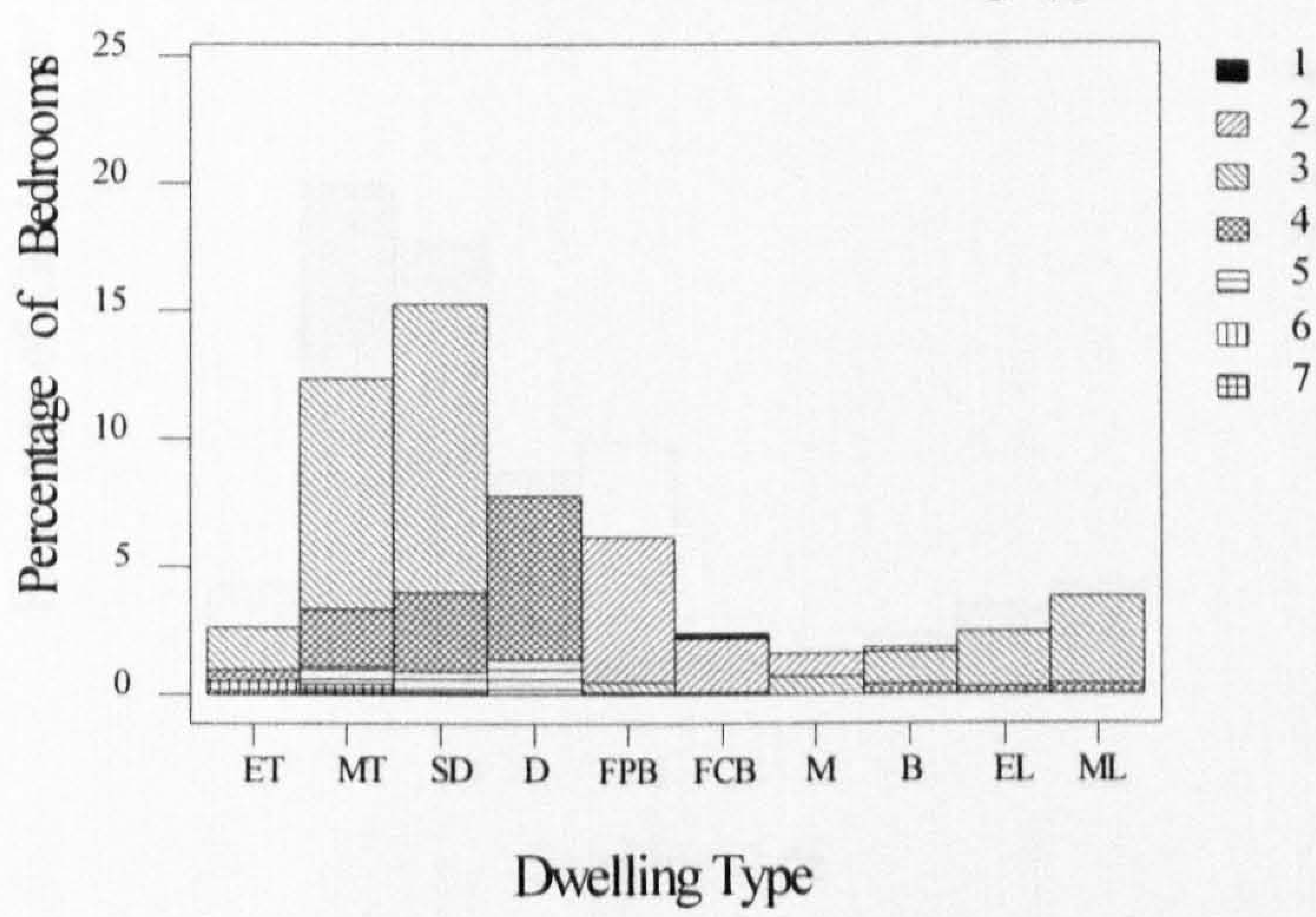


Figure 6.9
Number of Recreation Rooms in each Dwelling Type

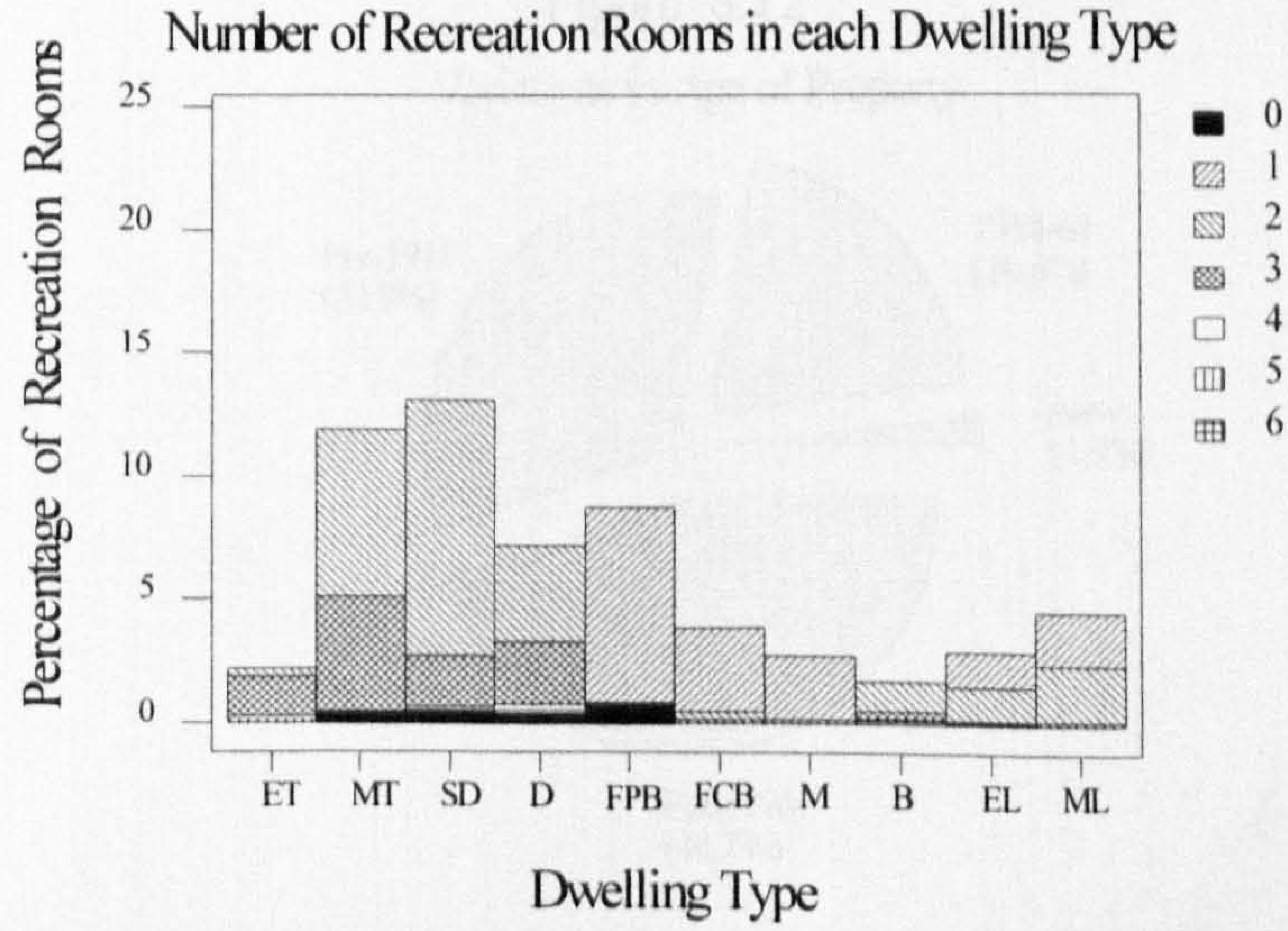


Figure 6.10
Variations in Garden Size

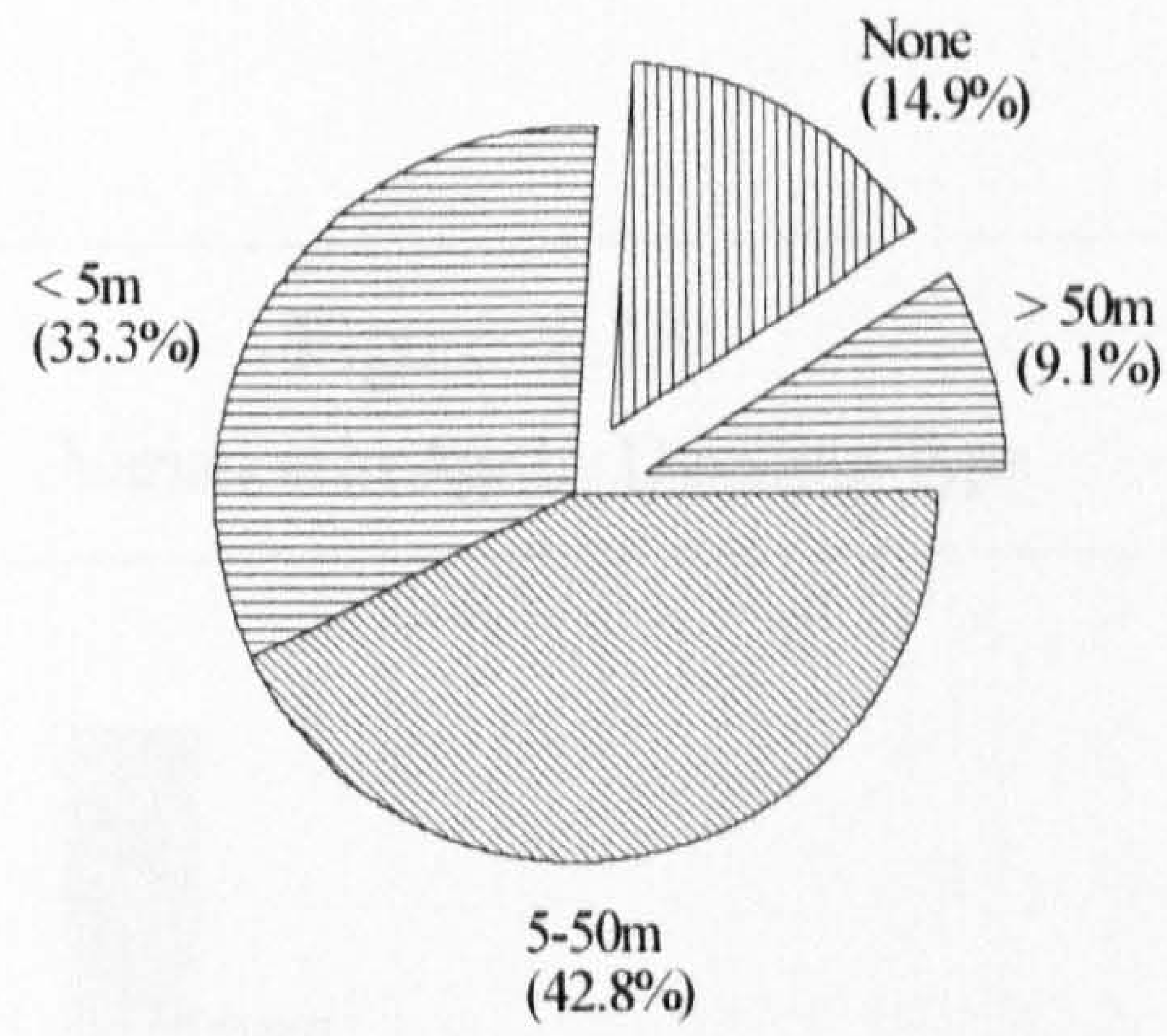


Figure 6.11
Variation in Garden Size by Dwelling Type

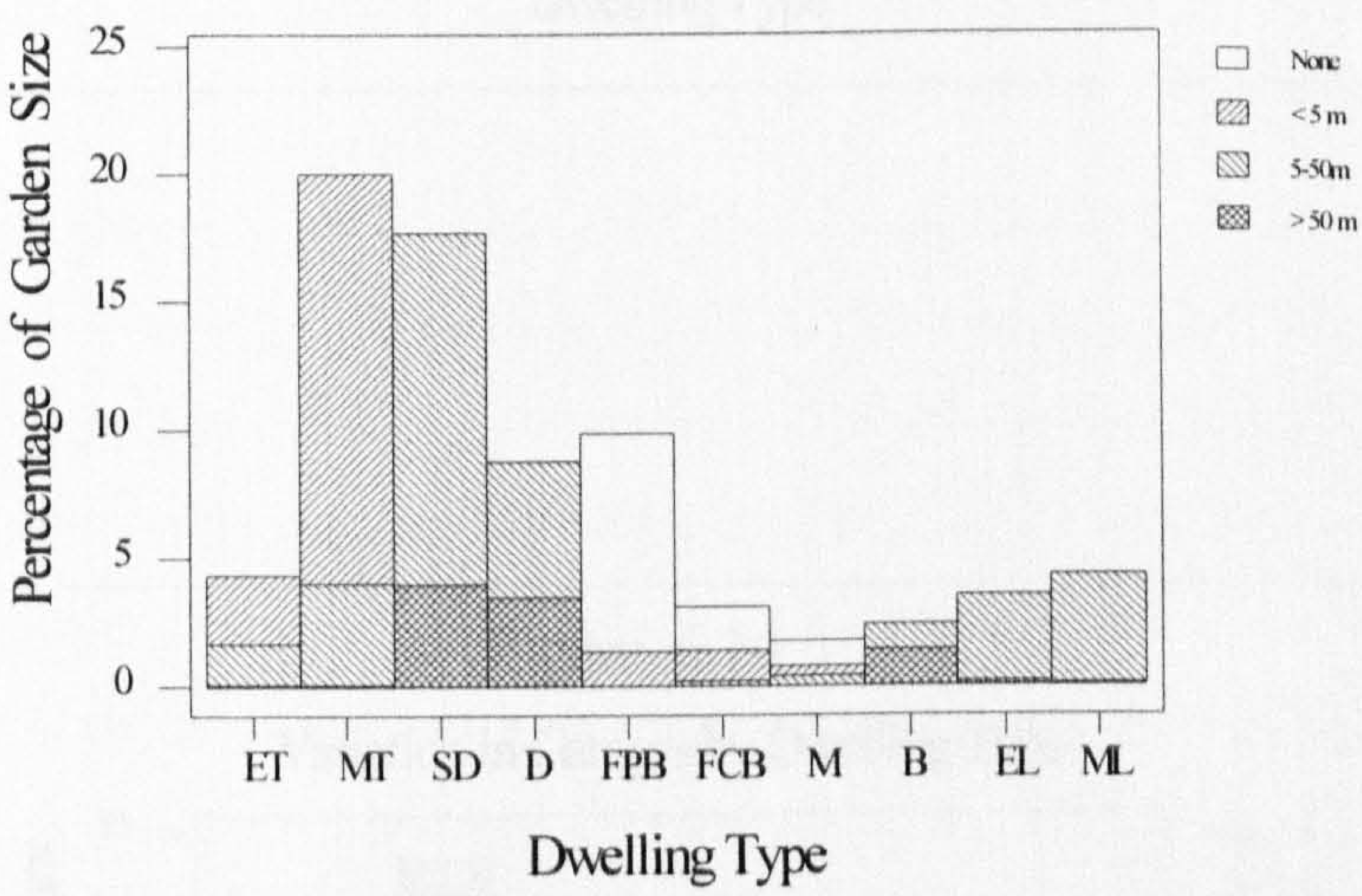


Figure 6.12
Variations in Age of Property

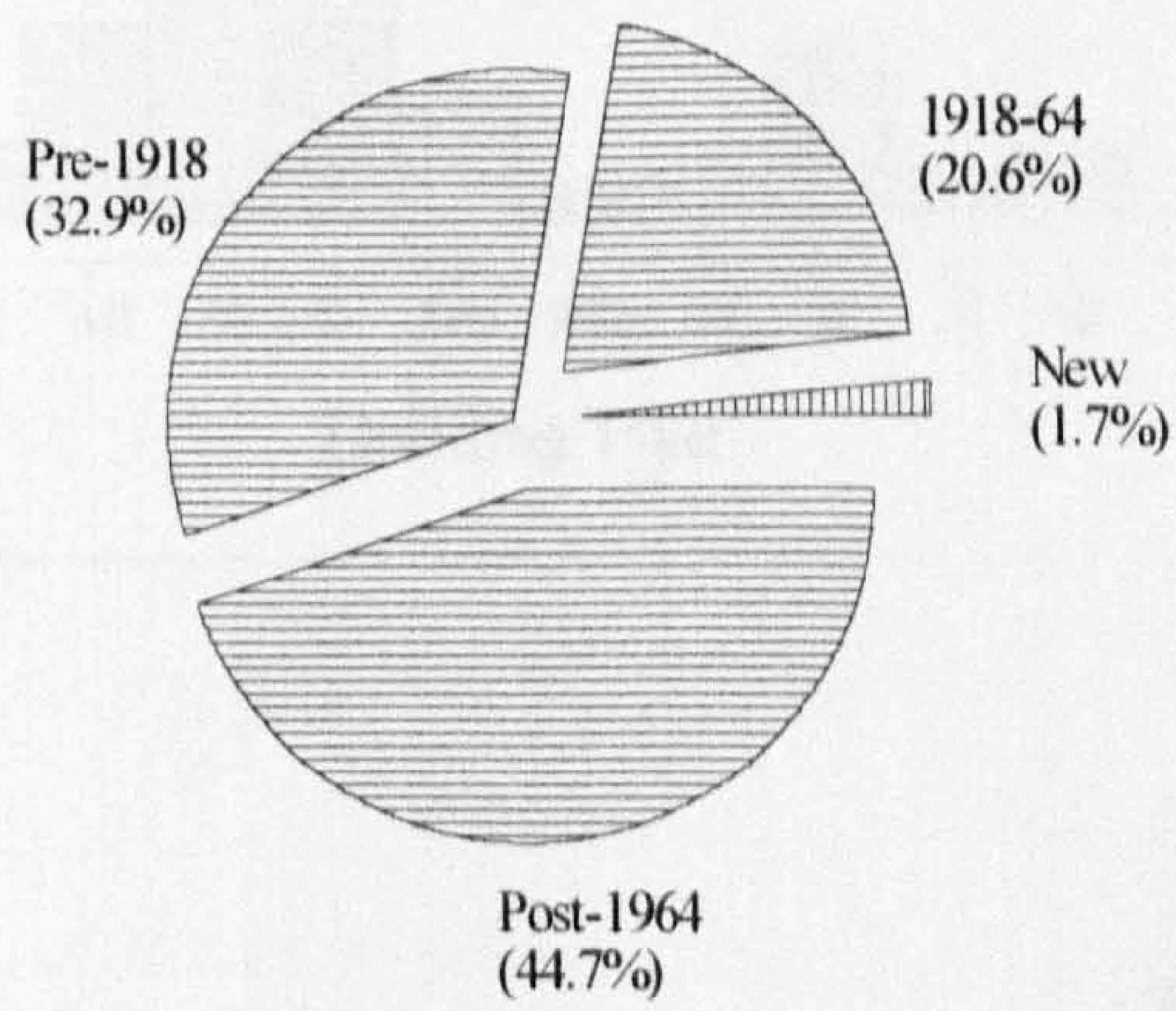


Figure 6.13

Variation in Age by Dwelling Type

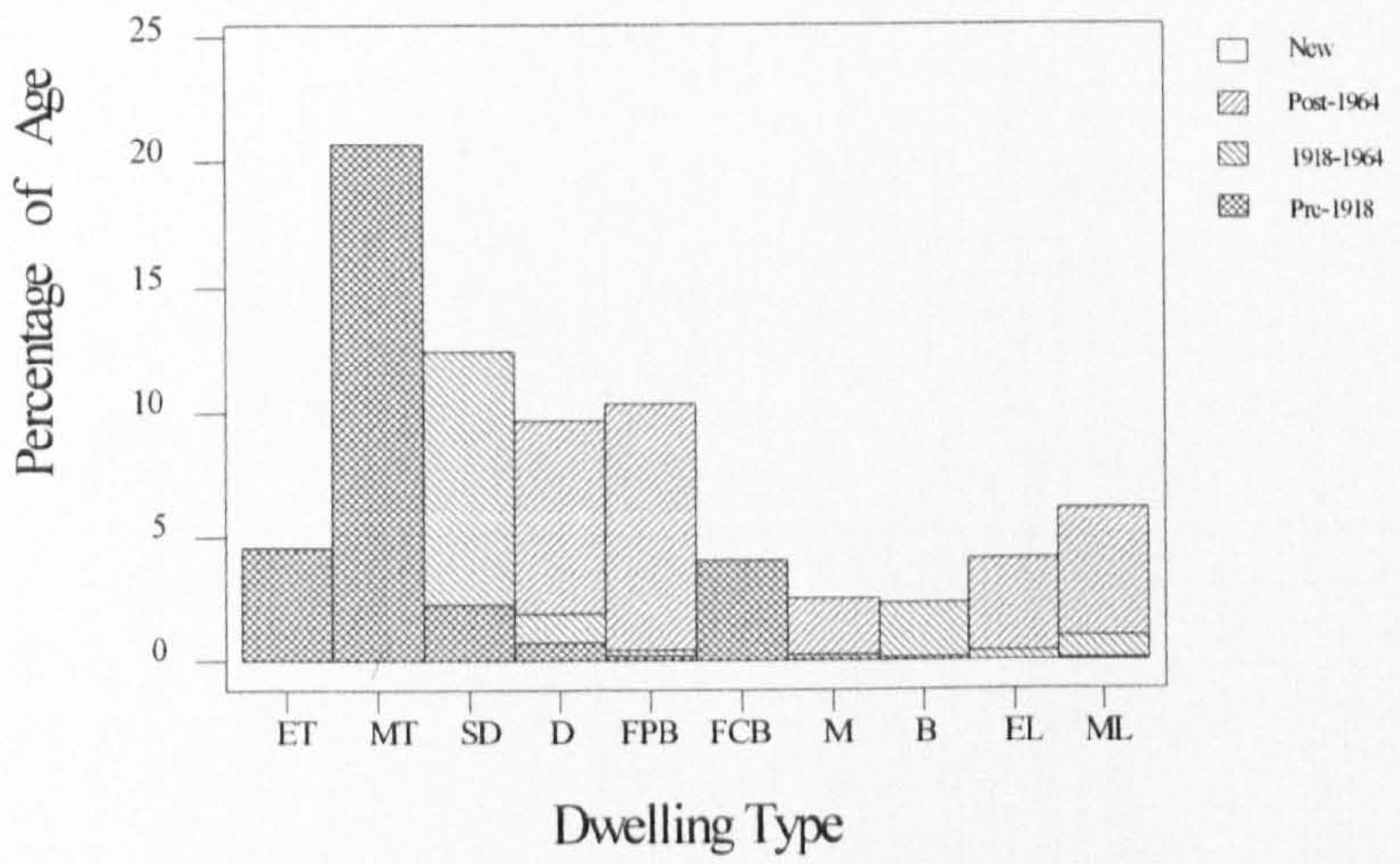
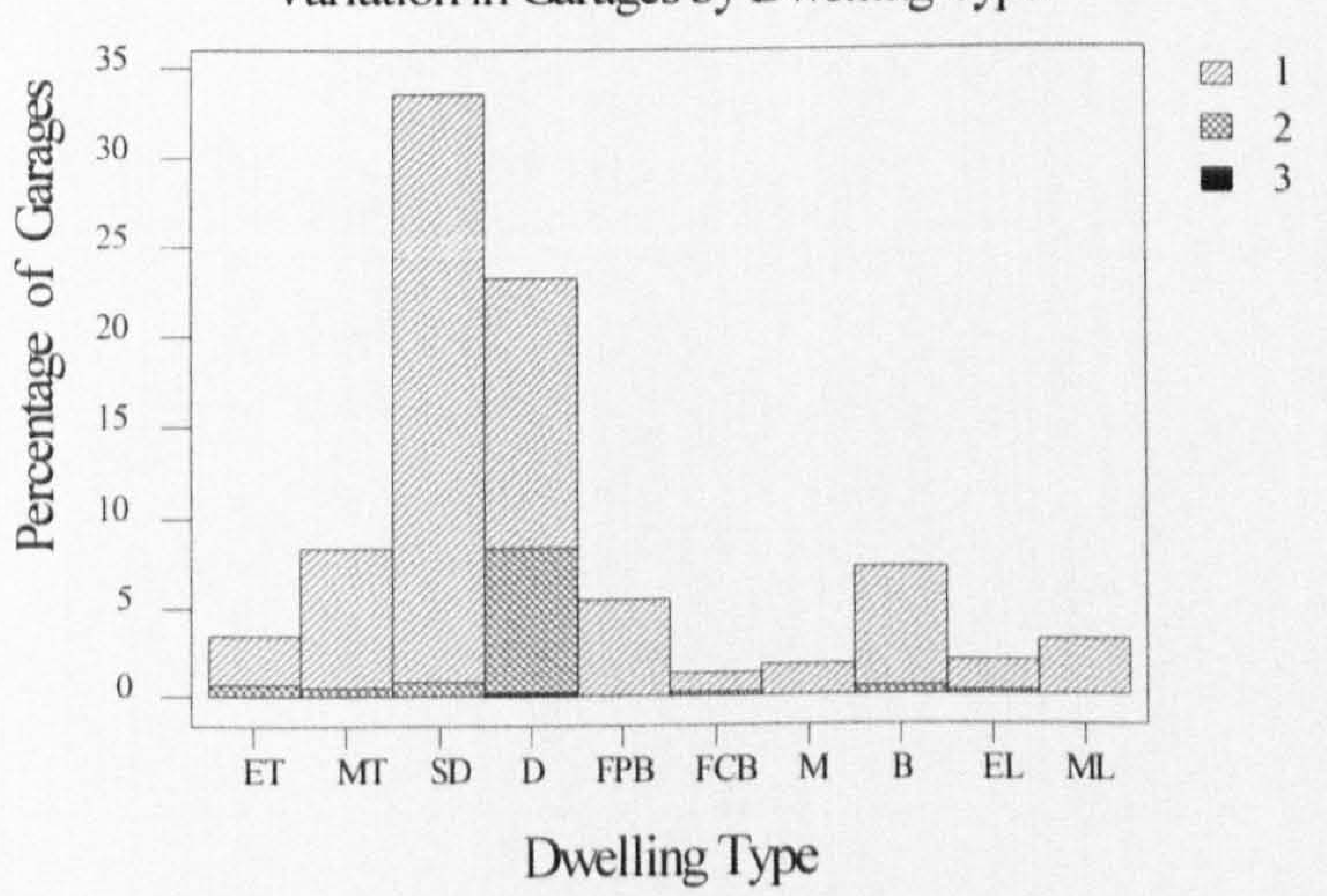


Figure 6.14

Variation in Garages by Dwelling Type



that these are generally flats, whilst houses with gardens in excess of 50 metres are confined to detached and semi-detached properties. A similar distribution describes the age of properties - Figure 6.12 - with a third having been built before 1918, and almost a half after 1964. The former constitute all the terraced houses and the flats in converted buildings, whilst the latter are typified by linked properties, purpose built flats and maisonettes - Figure 6.13. The majority of semi-detached houses and bungalows were constructed in the intervening years. The sample only contained a very small number of newly constructed properties, reflecting both the depressed state of the housing market at the time of the survey, and the fact that there are few sites available for development with the built up area. Figure 6.14 summarises the distribution of garages between dwelling types. A third of all the one car garages belong to semi-detached houses, whilst the majority of two car garages and all three car garages belong to detached properties. The majority of the remaining dwelling types, with the exception of bungalows, have no garage facilities.

6.2.2 A Comparison Against the Cardiff Housing Condition Survey

It was previously discussed in *Chapter Five* that 124 properties (c. 20%) sampled in the house price survey were also sampled in the CHCS. This overlap will allow a comparison of the structural attribute data obtained during the house price survey of the Inner Area to be made against similar data collected during the CHCS. Such a comparison will allow a measure of how the somewhat subjective categorisations in the house price survey are consistent with those in the CHCS. Moreover, the comparison can be used to gauge the extent to which the 'needs modernising' dummy variable in the house price survey reflects the need and costs of structural improvements calculated in the CHCS. However, it is acknowledged that some variation is to be expected due to the time lapse between the two surveys.

As was explained in *Chapter Four,* five variables were common to both surveys: dwelling type, garden size, age, central heating and parking facilities. Unfortunately, attributes such as room size and floor area were not recorded in sufficient detail in the CHCS to warrant a comparison. Figures 6.15 - 6.20 are a summary of the differences between the data sets. The main discrepancies in the classification of dwelling type understandably occurred between flats in converted buildings (FCB), and purpose built flats and terraces, the latter explained by the fact that the majority of converted flats are in terraced structures. The principal difference in garden size occurred with respect to properties with a garden less than 5 metres

Figure 6.15

Survey Differences in Dwelling Type

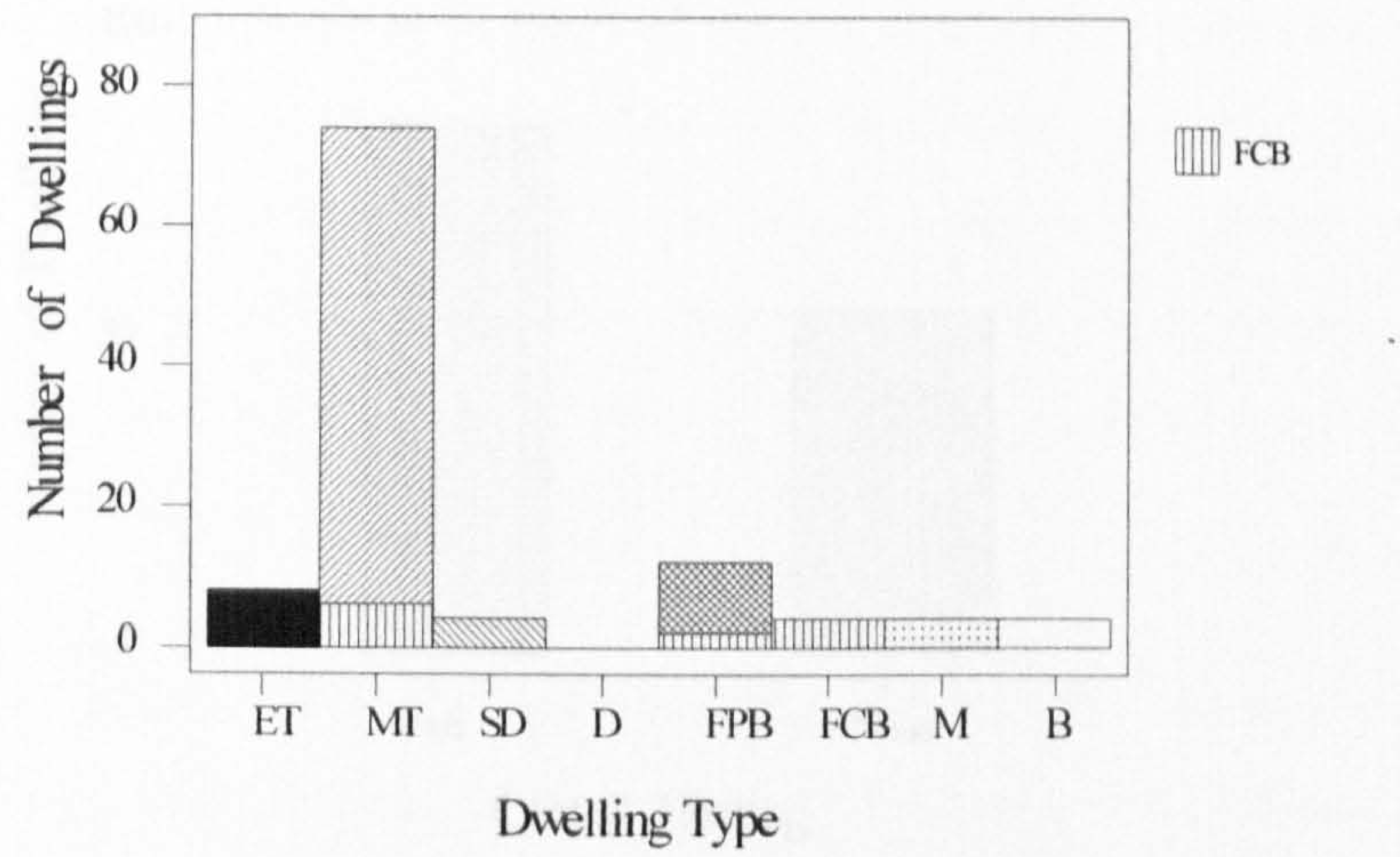


Figure 6.16

Survey Differences in Garden Size Categories

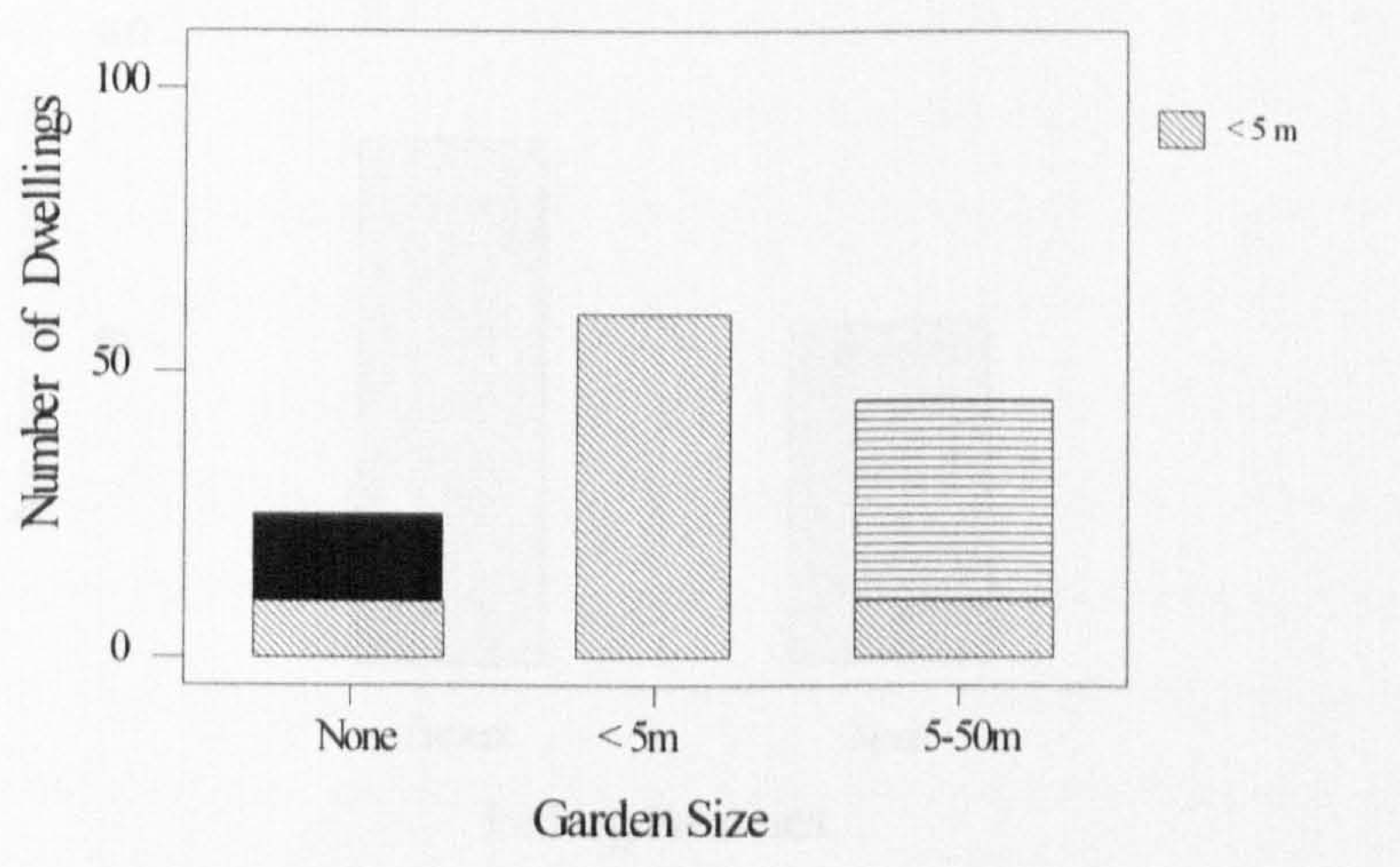


Figure 6.17

Survey Differences in Age Categories

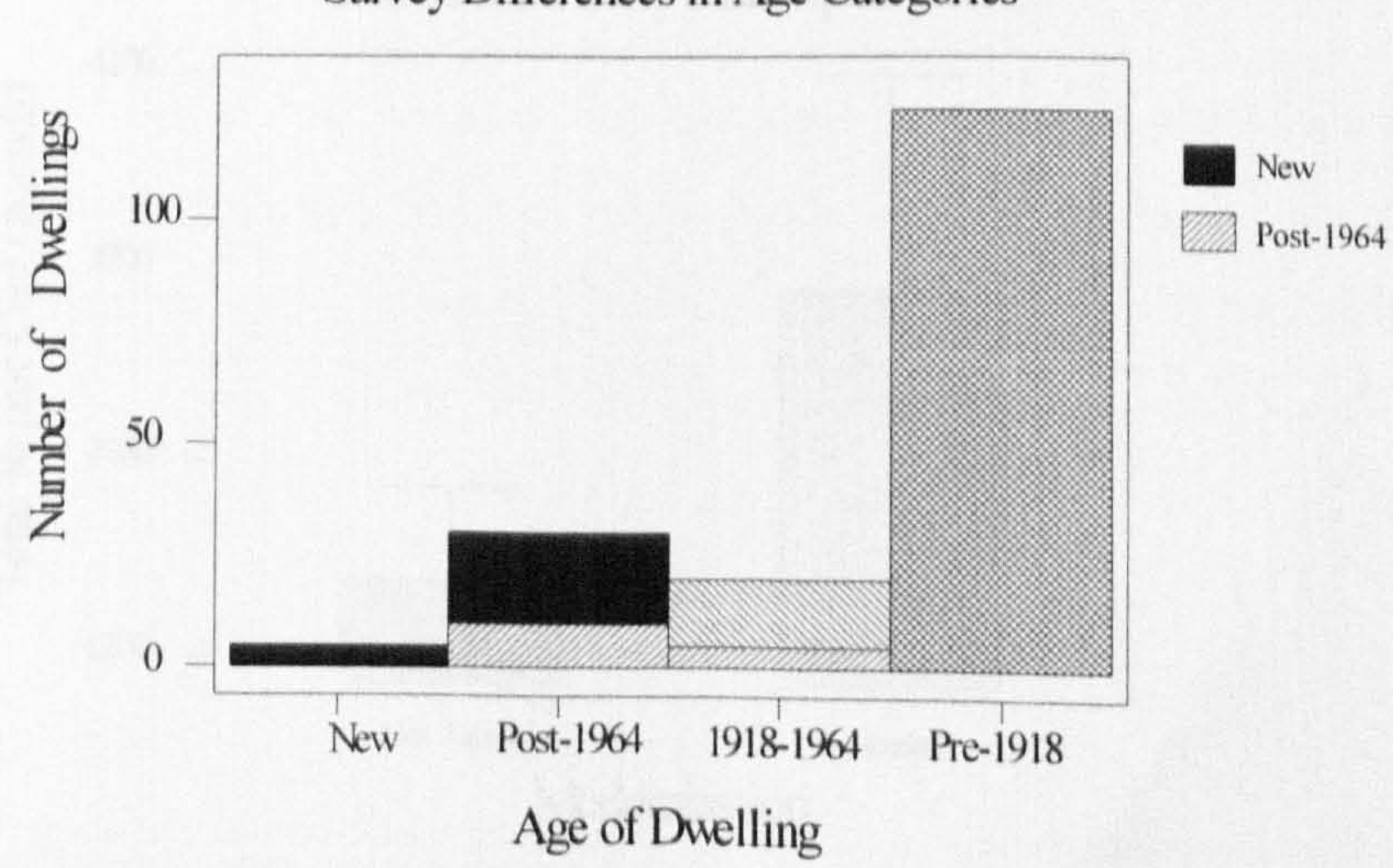


Figure 6.18

Survey Differences in Central Heating Categories

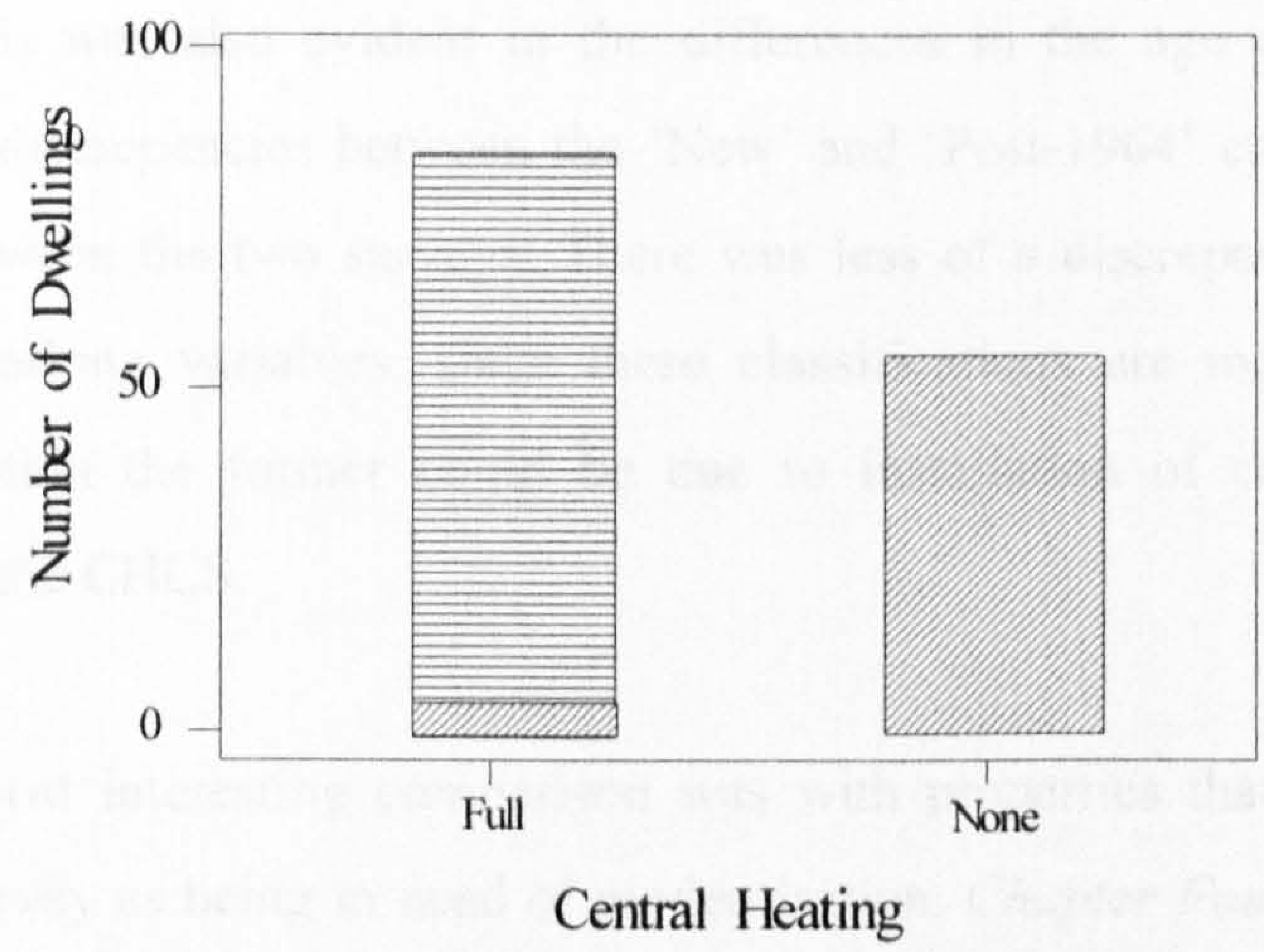


Figure 6.19

Survey Differences in Parking Categories

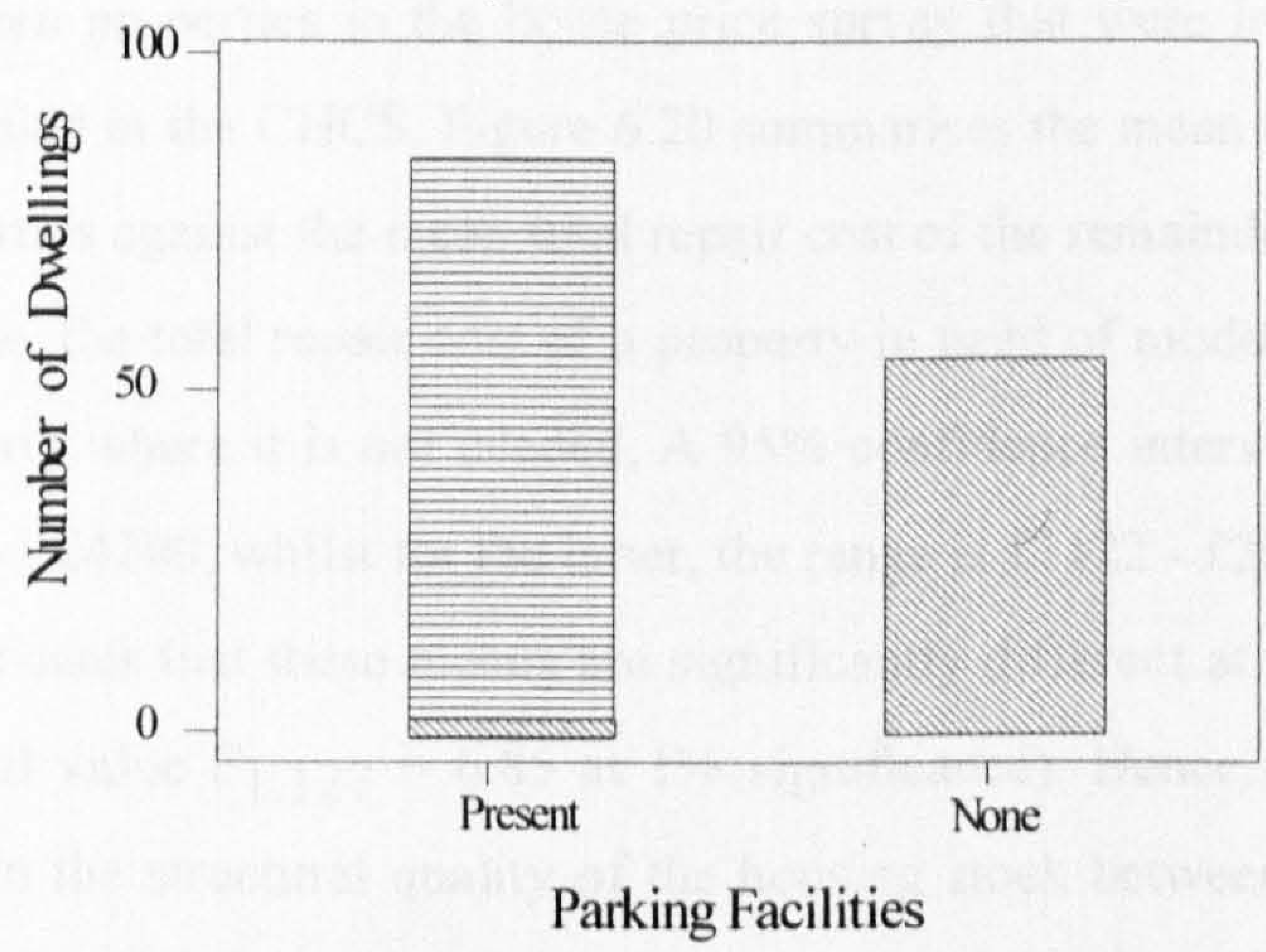
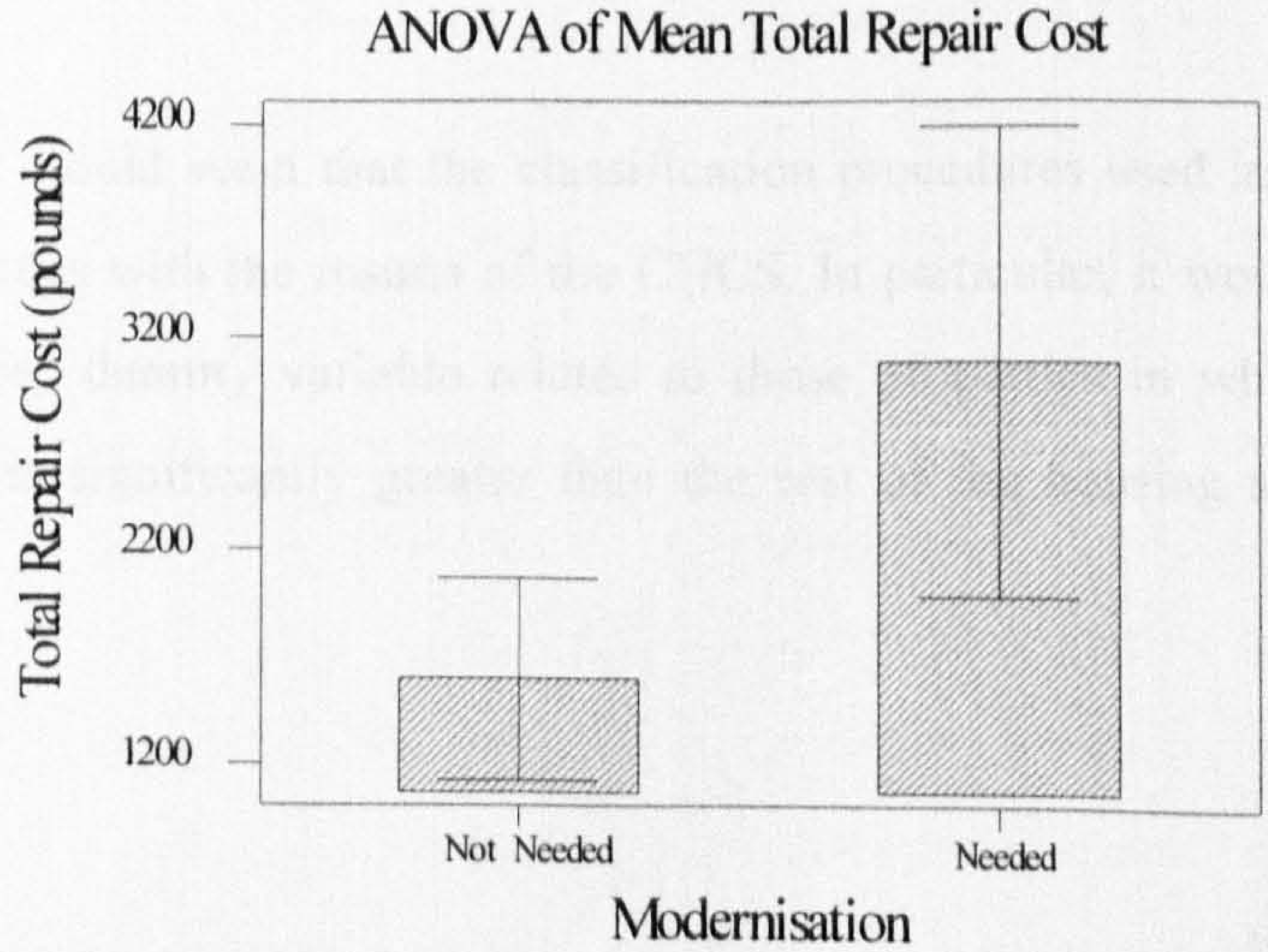


Figure 6.20

ANOVA of Mean Total Repair Cost



in length. The CHCS placed around 10% of these properties into the 'none' garden category, and 10% in the '5-50 metre' category, reflecting the arbitrariness of the classification procedure. This was also evident in the differences in the age classifications, although obviously the discrepancies between the 'New' and 'Post-1964' categories was due to the time lapse between the two surveys. There was less of a discrepancy between the central heating and parking variables, since these classifications are more tangible. Moreover, differences within the former could be due to installation of central heating after the completion of the CHCS.

Perhaps the most interesting comparison was with properties that were identified in the house price survey as being in need of modernisation. *Chapter Four* explained how a 'total repair cost' was estimated for each property in the CHCS, and this used to direct Local Authority improvement grants. By examining how this total repair cost estimate varies with respect to properties in the house price survey, the 'needs modernisation' variable can be gauged. Fourteen properties in the house price survey that were in need of modernisation were also sampled in the CHCS. Figure 6.20 summarises the mean total repair cost of these fourteen properties against the mean total repair cost of the remainder. It can be clearly seen that, on average, the total repair cost of a property in need of modernisation is nearly twice that of a property where it is not needed. A 95% confidence interval for the former ranges between £1906 - £4200, whilst for the latter, the range is £1122 - £2082. A one-way analysis of variance indicates that these means are significantly different at the 1% level (F-statistic = 11.67, critical value $F_{1,127}$ = 6.85 at 1% significance). Hence, taking into account the likely change in the structural quality of the housing stock between the estimation of total repair cost in the CHCS, and the house price survey, it would appear that the 'needs modernising' variable does indeed reflect properties in a significantly greater state of disrepair.

Therefore, it would seem that the classification procedures used in the house price survey generally agrees with the results of the CHCS. In particular, it would seem that the 'needs modernisation' dummy variable relates to those properties in which the estimated 'total repair cost' is significantly greater than the rest of the housing stock at the time of the CHCS.

Section 6.3 GIS and the Locational Attribute Data.

6.3.1 Introduction

One of the primary purposes of the GIS has been to generate locational specific attribute data. Such data will allow the exploration and evaluation of the two concepts of locational externalities: that their effect upon property values diminishes with distance, and that they operate over different spatial scales. The construction of the locational attribute data has been a two-fold process. Firstly, attribute data have been generated for the whole of the Cardiff, although in effect, the use of the GIS in this process was minimal. Instead, the GIS has been extensively used in generating detailed locational attribute data for the Inner Area. Hence, this section is structured into two parts, with the emphasis upon GIS in the second section.

6.3.2 Locational Attribute Data for the Cardiff Housing Market

Since the purpose of the Cardiff housing market study is to explore spatial housing market dynamics replicate within conventional urban theory, traditional measures of location were used. This involved the construction of a measure of accessibility to the city centre and measures of environmental quality calculated from Census data at the Enumeration District level.

6.3.2.1 Accessibility Measures

A standard function of any GIS has been to calculate accessibility between points of interest. In the macro-study, accessibility was reduced to a simple measure of straight line distance from each property to Cardiff city centre. This was calculated within ARC / INFO using the POINTDISTANCE command which calculated the distance between the city centre centroid to all the points in the macro-scale property coverage. Although this can be quite coarse due to the problems associated with postcoded grid-references, it is quite accurate when compared to accessibility measures used in previous studies. To ameliorate the problems of estimating the hypothesized rent gradient when the functional form is not known, the linear accessibility measures were divided into distance intervals. This has the advantage that the rent gradient is not constrained by any of the *a priori* imposed functional forms suggested by previous studies. Furthermore, as was argued in *Chapter Three*, the presence of local maxima and minima may result in the rent gradient not fitting perfectly into any of these generalised functional forms, but fluctuating along a downward trend. These fluctuations may be captured by the distance intervals.

Table 6.2 is a summary of the distance intervals used. The intervals used were based upon the accepted theory that rent gradients are generally non-linear and decrease at a decreasing rate from the city centre. Hence, the curve will be at its steepest within the first couple of

Table 6.2

Distance Intervals from Cardiff City Centre

Distance (metres)	Sample Size	Distance (metres)	Sample Size	Distance (metres)	Sample Size	Distance (metres)	Sample Size
0-100	6	1000-1100	44	2000-2200	27	4000-4200	54
100-200	7	1100-1200	46	2200-2400	27	4200-4400	59
200-300	11	1200-1300	49	2400-2600	14	4400-4600	35
300-400	10	1300-1400	36	2600-2800	24	4600-4800	52
400-500	27	1400-1500	31	2800-3000	26	4800-5000	57
500-600	33	1500-1600	35	3000-3200	44	5000-5500	89
600-700	43	1600-1700	44	3200-3400	25	5500-6000	100
700-800	35	1700-1800	19	3400-3600	41	6000-6500	69
800-900	25	1800-1900	26	3600-3800	23	6500-7000	55
900-1000	27	1900-2000	20	3800-4000	42	7000-9000	43

kilometres, after which the curve will eventually flatten. This is reflected in the selection of distance intervals, with 100 m intervals for the first two kilometres to capture the rapid change, then 200 m intervals for the next three kilometres followed by 500 m intervals for the remaining two kilometres where the curve is hypothesized to be gentle with very little variation. The study area extends a further two kilometres, and this broad interval was chosen as the final distance measure, reflecting the lack of sampled properties at the rural fringe.

6.3.2.2 Locational Quality Measures

Three locational attributes measuring environmental quality were constructed from the Census data described in *Chapter Four*. Table 6.3 summaries the variables constructed from these data. Because of their low representation in the housing stock, the four variables that measured the lack of basic amenities (the percentage of households with no, or a shared, bath or shower and no, or shared, inside WC) were aggregated into one variable that

Table 6.4
The Correlation Matrix of the Census Variables

	M.U e	F.Ue	N.White	Amenities	CH	No car	Two Car	Lone	Young	Pensioner	Families
F.Ue	0.241										
N.White	**0.437**	0.108									
Amenities	0.232	0.298	**0.409**								
CH	0.357	0.399	0.382	**0.694**							
No car	**0.847**	-0.142	**0.402**	0.297	0.528						
Two car	**-0.709**	0.221	-0.335	-0.253	**-0.497**	**-0.875**					
Lone	**0.630**	0.323	0.209	0.027	0.155	**0.540**	**-0.441**				
Young	**0.527**	**-0.483**	0.326	**0.433**	**0.503**	**0.521**	**-0.648**	**0.427**			
Pensioner	**-0.876**	0.09	-0.374	**0.664**	**0.455**	**0.558**	**-0.447**	-0.317	0.225		
Families	**0.521**	0.362	**-0.592**	-0.238	**-0.551**	-0.325	0.397	-0.274	-0.374	**-0.455**	
Owner	**-0.823**	-0.187	-0.389	-0.326	-0.287	**-0.824**	**0.663**	**-0.638**	-0.359	**0.566**	**0.592**

Key to variable abbreviations

M.Ue	Percentage of male unemployment	Two car	Percentage of households with two or more cars
F.Ue	Percentage of female unemployment	Lone	Percentage of lone parent households
N.White	Percentage of non-white households	Young	Percentage of households young and single
Amenities	Percentage of households which have no, or share, a bath or shower or WC	Pensioner	Percentage of households of pensionable age
CH	Percentage of households with no central heating	Families	Percentage of households married with family
No car	Percentage of households with no car	Owner	Percentage of households in owner occupied tenure

measured the overall lack of basic amenities. A correlation matrix (Table 6.4) of the variables indicated a high degree of collinearity, notably between the percentage of male unemployment, the household structure variables and the number of cars. For this reason, and because individually, the variables will only marginally capture environmental quality, principal components analysis was used to construct new indices that would proxy locational attributes more effectively. In addition, the 'Percentage of households in Local Authority tenure' variable was treated separately, and used to construct a measure to capture the 'stigma' of council built housing stock. This is explained in detail in the next section

Table 6.3

Variables Constructed from 1991 Census Data

Socio-economic dimension	Percentage of male unemployment Percentage of female unemployment Percentage of lone parent households Percentage of households with no car Percentage of households with two or more cars Percentage of households with no, or a shared bath or shower or inside WC Percentage of households with no central heating Percentage of households in owner occupied tenure Percentage of households in Local Authority tenure
Family life-cycle dimension	Percentage of households young and single Percentage of households pensioners Percentage of households married with family
Ethnic dimension	Percentage of non-white households

6.3.2.3 Principal Components Analysis

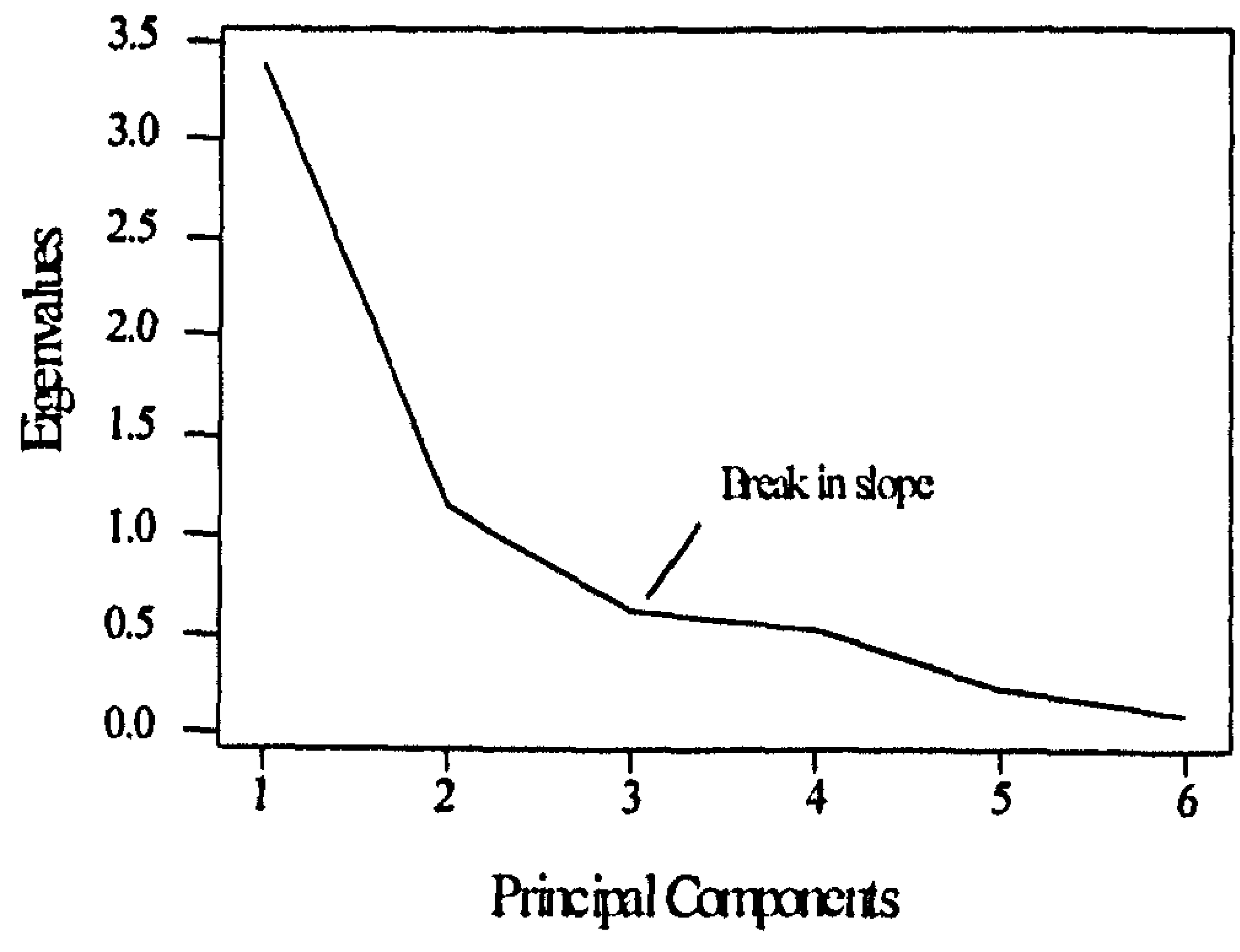
Principal components analysis is a synthesizing technique that is able to identify groups of variables that have similar patterns of variation. Once identified, these groups of variables can be transformed into hybrid variables called components which summarise the original data. The process produces a series of uncorrelated components, each accounting for a successively smaller amount of the co-variance between the original variables. Hence, the technique allows variables that have a high degree of multicollinearity to be transformed into new, uncorrelated components. It is then common for these components to be used as the inputs to subsequent analysis.

Table 6.5
Summary of the Principal Components Analysis

Eigenvalue	3.387	1.146	0.626	0.526	0.231	0.084
Proportion of explained variance	0.564	0.191	0.104	0.088	0.039	0.014
Cumulative explained variance	0.564	0.755	0.860	0.947	0.986	1.000

Variables	Component 1	Component 2	Component 3
Percentage male unemployment	-0.494	-0.144	0.056
Percentage female unemployment	-0.254	-0.051	0.032
Percentage households non-white	-0.312	0.509	0.655
Percentage households shared / no amenities	-0.509	0.728	-0.277
Percentage of households with no central heating	-0.556	0.486	-3.71
Percentage households owner occupied tenure	0.501	-0.098	-0.269
Percentage households no car	-0.509	0.104	-0.269
Percentage households two or more cars	0.468	0.104	0.440
Percentage households young/single	-0.402	0.517	0.176
Percentage households old	-0.338	0.445	-0.214
Percentage households families	0.214	0.179	-0.165
Percentage households lone parent family	-0.363	0.475	0.475

Figure 6.21
Plot of Eigenvalues against Components



With the exception of the 'Percentage of households in Local Authority tenure', the variables in Table 6.3 underwent principal components analysis, the output of which is summarised in Table 6.5. The first part of the table represents the eigenvalues of the correlation matrix, together with individual and cumulative percentages of the total variance. This suggests that three quarters of the variation within the data lies in two-dimensional space, whilst 86 % lies in three-dimensional space. Higher dimensions each contain only a negligible proportion of the total variability. Since the objective is to identify only the major dimensions of co-variance in the data, it is usual to retain only those components which account for a greater proportion of the total variance than could any of the original variables.

Therefore, the number of principal components that summarise the data lies somewhere in the range of two to three. Several methods have been suggested to determine the optimum number of components that best describe the data (Mather, 1976). One method uses the 'scree test', in which the eigenvalues are plotted against the number of components, and the break in slope indicates the point of discrimination between the useful and trivial components (Figure 6.21). Another procedure argues that since each variable has a variance of one when expressed in standard form, any principal component with an eigenvalue less than one is not worth consideration. Both methods were considered. It was decided that two principle components were satisfactory in describing the data on the grounds that the first two components accounted for three quarters of the total variation, that they both had eigenvalues in excess of one, and the two components were interpretable with respect to the original data. This latter point shall now be discussed.

6.3.2.4 Interpretation of the Two Components

The second part of Table 6.5 summarises the loadings of each component. These represent correlations between the principle components and the original data, and indicate how much a variable has contributed to the construction of a particular component. In the first component, the range of the magnitudes of the loadings are quite small (0.214 - 0.556), suggesting that several variables have had an equal contribution. However, the highest loadings are associated with indicators of income. It is negatively correlated with areas of poor housing conditions, high male unemployment, single / young persons households, and a low degree of car ownership. Conversely, it is positively correlated with areas of owner occupation and two car households. A map of component one (Figure 6.22) reveals that the

Figure 6.22

The Geography of Component One (Socio-Economic Class)

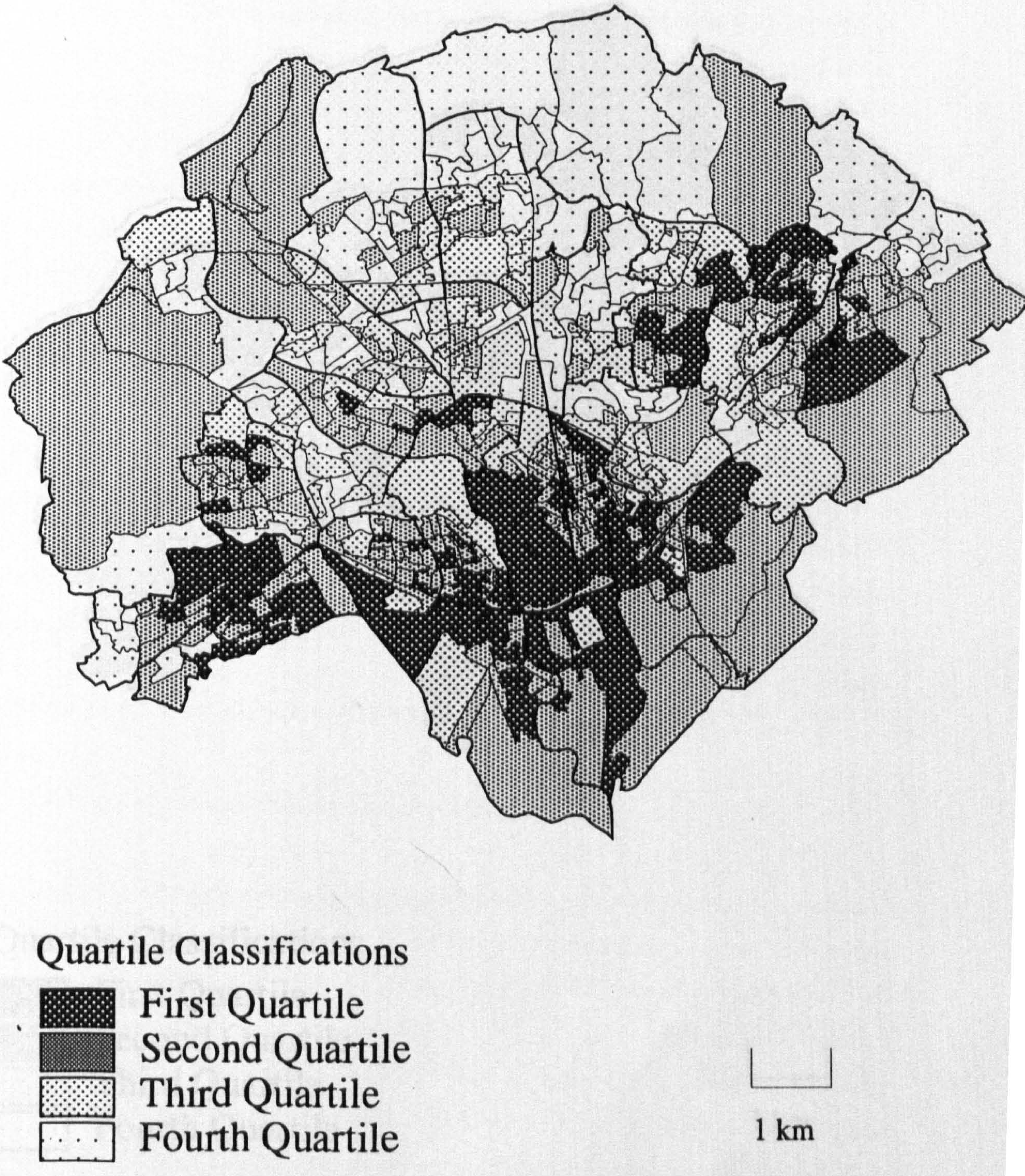
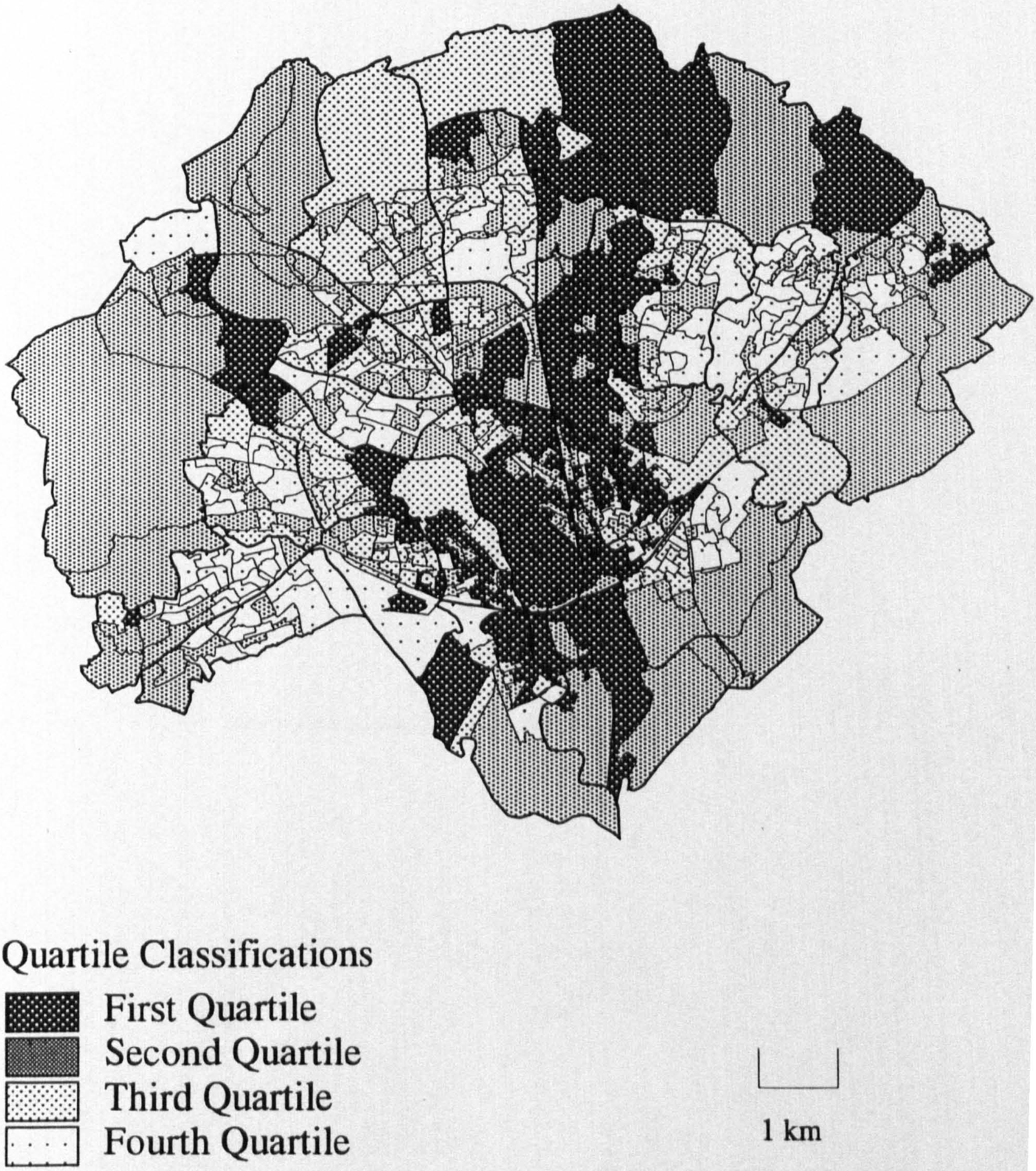


Figure 6.23

The Geography of Component Two (Housing Quality)



lower quartile of the data is clustered in the dockland areas of Butetown and Grangetown, and the rundown inner city neighbourhoods of Riverside, and Adamsdown, and also in the council estates of Ely, Caerau, Pentwyn and Tremorpha. A concentration is also located in the peripheral estate of Trowbridge. The upper quartile of the data is clustered in suburbs to the north of Cardiff such as Llanishen, Lisvane and St Mellons and Rhiwbina, and also in Roath and Cyncoed. Hence, this geography and the above loadings suggests that component one can be classified as a general measure of socio-economic class.

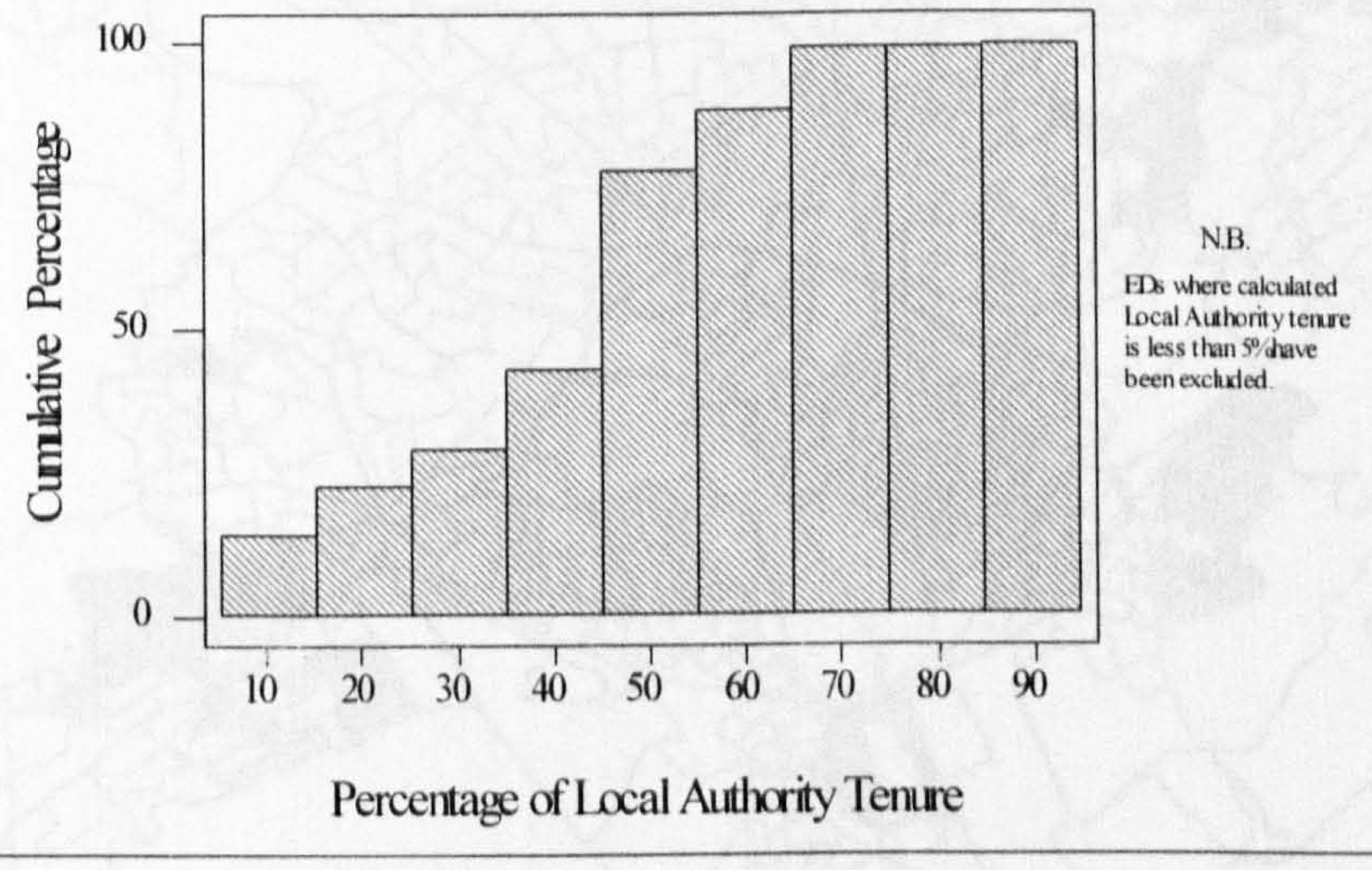
The loading in the second principal component can be allotted into three groups of significant variables. The highest loading identifies areas of housing in poor condition as being significant. Next are areas that have a high percentage of non-white households and young / single households. Finally are areas of that have a high percentage of elderly, and single parent households. It can be seen that unlike the former, these loadings are quite diverse. However, inspection of the data suggests that they can all be associated with areas of poor housing condition, the first loading directly so. This is qualified by a map of component two - Figure 6.23. Areas of strong, positive association are located in the peripheral council estates, and the inner-city neighbourhoods to the east of the city centre. The housing condition survey concluded that these neighbourhoods contained a high concentration of houses in need of modernisation. Areas of negative association are located in inner city neighbourhoods that have undergone a process of renewal or modernised, such as parts of Grangetown, Riverside and Plasnewydd, and also the suburbs to the north of Cardiff. However, these suburban areas contain pockets of high, positive values that can be related to areas of elderly households. Hence, although the explanation of the component is more vague compared to the previous, it can be concluded that it is generally associated with areas of housing in a relatively poor condition than Cardiff as a whole.

6.3.2.5 Areas of Local Authority Housing Stock

The percentage of households in Local Authority tenure was used to construct a dummy variable representing those EDs which had over 50 % of their housing stock in Local Authority ownership. This attribute reflects the stigma, acknowledged in literature and by estate agents, that affects areas of housing stock of Local Authority origin. This was also concluded in the study of local taxation, in which Local Authority built properties gained in the switch from rateable value (use value orientated) to the council tax (exchange value orientated), suggesting that the externalities associated with Local Authority built housing estates had a negative effect upon property prices (Longley et al, 1993). The 50 % cut off mark was chosen since analysis of the data suggested that EDs contained either a very high or very low percentage of Local Authority owned housing, and a 50 % threshold represented a meaningful break - see Figure 6.24. Also, it was assumed that such stigma would only be perceptible in areas where houses of Local Authority origin dominated the housing stock. One problem with this approach it that it does not take into account Right to Buy properties. Areas of Local Authority built housing could exist in Cardiff that are now predominately owner occupied, but still have a stigma attached to them because of their origins. Since Right to Buy properties are not distinguished in the 1991 Census, EDs that have a predominance of Right to Buy properties may not be flagged by the dummy variable, even though over fifty percent of the housing stock may be of Local Authority origin. This problem is difficult to rectify without recourse to extensive field work, which is beyond the scope of this study. Figure 6.25 illustrates those EDs where the stigma affect would be expected to be most prominent.

Figure 6.24

Distribution of Local Authority Tenure amongst EDs



6.3.3 The Cardiff Inner Area Study

It was argued in *Chapter Three*, section 3.4 that locational externalities could be hypothesized as operating across four spatial scales, representing the four basic structures of the Cardiff Inner Area: properties, streets, HCS areas and communities. Coverages pertaining to these spatial scales were constructed for the Inner Area in *Chapter Five*. Using the findings of *Chapter Five*, Table 6.6 is a summary of the externalities that are

Figure 6.25

EDs with Local Authority Tenure Greater than Fifty Percent

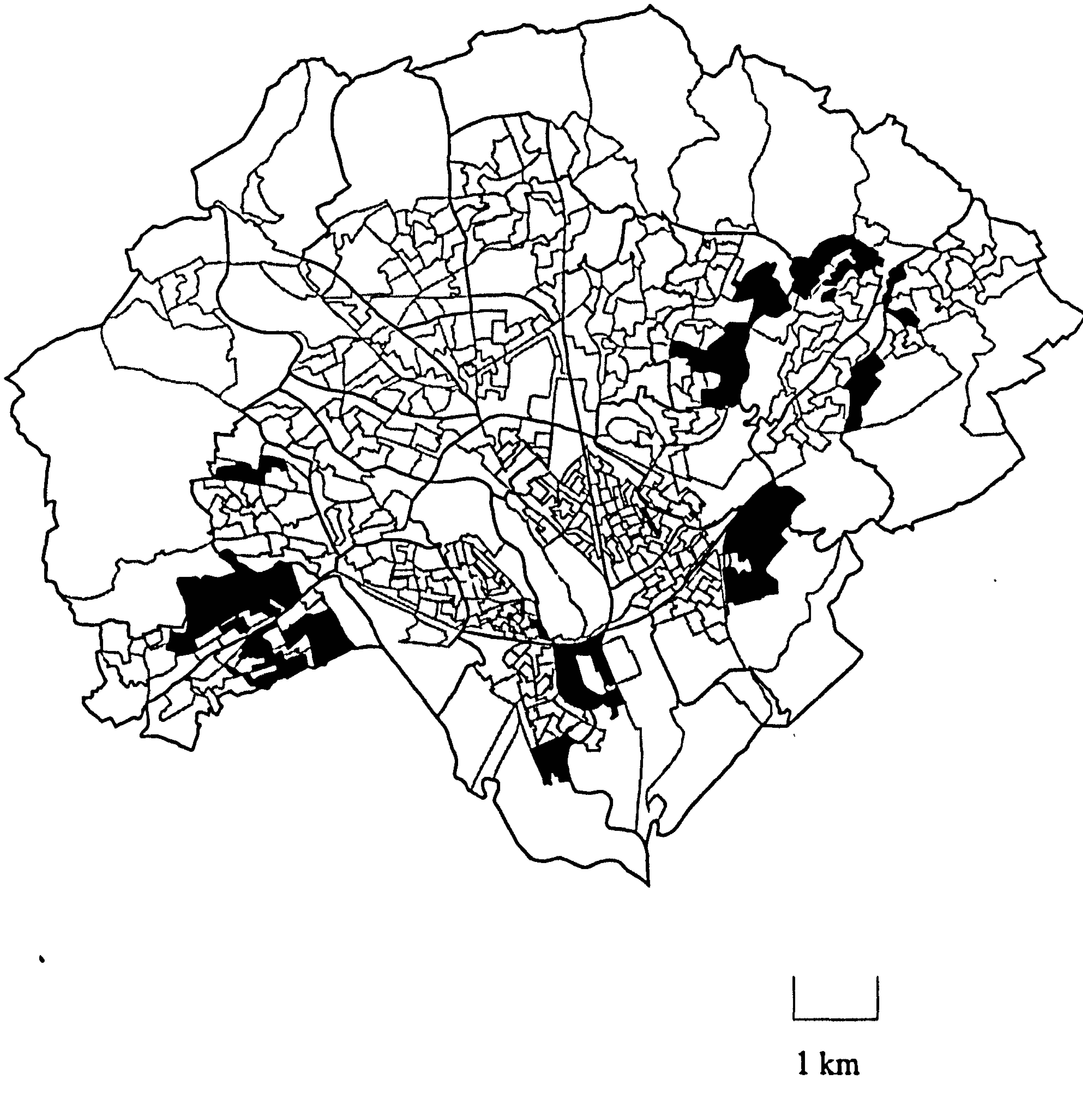


Table 6.6.
The Spatial Scale of Inner Area Locational Externalities

Property Level Attributes	Accessibility measures to work place City centre Motorway (M4) junctions Railway stations Proximity measures to non-residential landuses Hospitals Sports centres Community centres Local shops Primary schools Secondary schools Bute Park Parks / open space Light industry Heavy industry Institutional centres Railway Lines River Taff
Street Level Attributes	Street environment measures Class of street Street quality Non-residential activity School catchment areas
HCS Area Level Attributes	Percentage of open space Percentage of non-residential landuse Housing density Quality of local amenities Percentage of Local Authority tenure
Community Level Attributes	Social composition

hypothesized to operate at each of these four levels. This next section describes how measures of externality effects were generated within ARC / INFO at each of these spatial scales.

6.3.3.1 Property Level Externalities

Table 6.6 identifies the two sets of externalities that are hypothesized to operate at the property level. The first pertain to accessibility to place of work, whilst the second relates to proximity to non-residential landuse. Measures for both sets of externalities were generated using a variety ARC / INFO commands, which are explained in detail in the next sub-sections.

I. Measuring Accessibility Using the GIS

It is hypothesized that access to Cardiff city centre, access to the M4 motorway and access to the rail network will be significant determinants of Inner Area house prices. Hence, the following describes the computation of these three measures of accessibility within ARC / INFO.

In the Cardiff housing market study, there was no information about the transport network, and hence accessibility was computed as Euclidean distance, with no regard for the underlying typology. However, in the Inner Area, both the road and rail network for the Inner Area were available as ARC / INFO coverages. In this case, accessibility between two points is the shortest route on the network connecting them. An additional advantage of using a network to measure accessibility is that the shortest route need not be measured solely in terms of distance. Other costs, called impedance costs, can be used when calculating accessibility. Impedance costs take into account factors that may affect accessibility between two locations in an urban area, such as speed limits. Within ARC / INFO, accessibility along the street network was calculated using the NETWORK module. Impedance costs were assigned as estimated travel times from each street. This was calculated by dividing the distance of each street by its speed limit. The speed limit was calculated using the information on the class of street (primary, secondary, residential, cul-de-sac) that was recorded in the CHCS, and the calculated travel times were subsequently attached to the street network using the procedure outlined in *Chapter Five*. Roads that have a lower travel time and thus impedance cost will be favoured when calculating accessibility.

Hence, this will reflect the better accessibility that properties adjacent to main roads enjoy. Of course, this method of calculating travel time does not take into account factors such as congestion, but it is argued that perceived access to a main road will out weigh any perception of congestion when purchasing a property.

One problem with using the street network is that the calculated travel times are assigned to the street network, and not the attributes in property coverage. However, it is possible to match each property to the nearest node on the street network in ARC / INFO. This was achieved using the NEAR command, which matches the identifier of the nearest node in the street network coverage to each property in the point coverage. This identifier then acts as a link to which travel time data can be matched to the property coverage. To ensure that the distance between the property and the nearest node is as minimal as possible, the DENSIFYARC command was used. This command adds nodes to the street network at specified intervals along each arc. The interval used was five metres, since this was optimal with regard accuracy of the GIS and computer processing considerations.

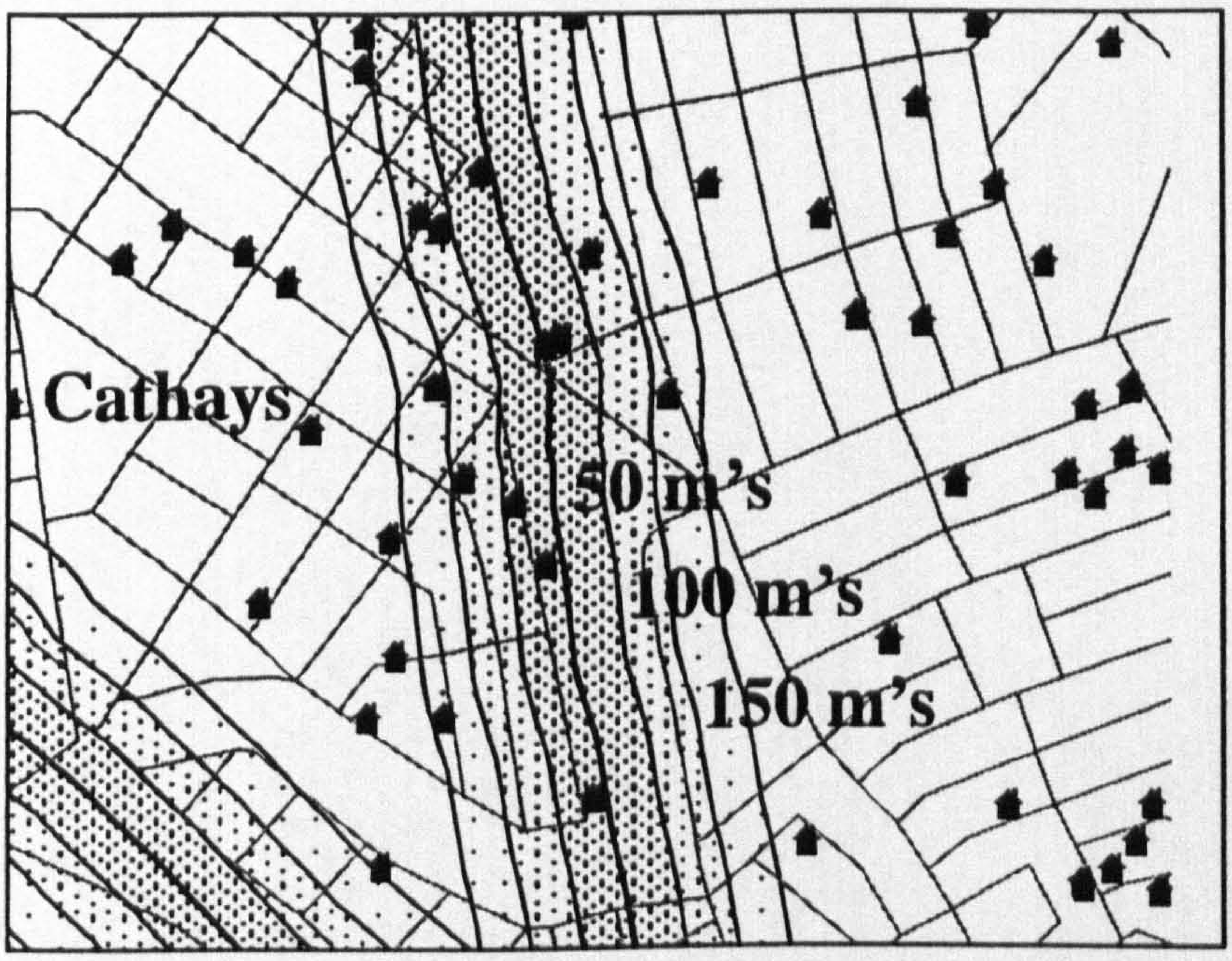
Accessibility to the city centre was calculated using the ALLOCATION command in the NETWORK module, working from the city centre outwards along the street network. This allowed the minimum travel time from the city centre to each node on the street network to be calculated simultaneously. Moreover, the ALLOCATION command allows more than one destination to be selected at any one time. Hence, accessibility to the nearest point of interest, such as the nearest railway station, can be calculated simultaneously for each property. Therefore, using this method, access to the CBD, the nearest of the four M4 motorway junctions and the nearest railway station in Cardiff were calculated separately for each property. A separate accessibility measure was also calculated for travel time to Cardiff Central railway station, the main rail terminus in the city.

II. Measuring Proximity to Non-residential Landuse

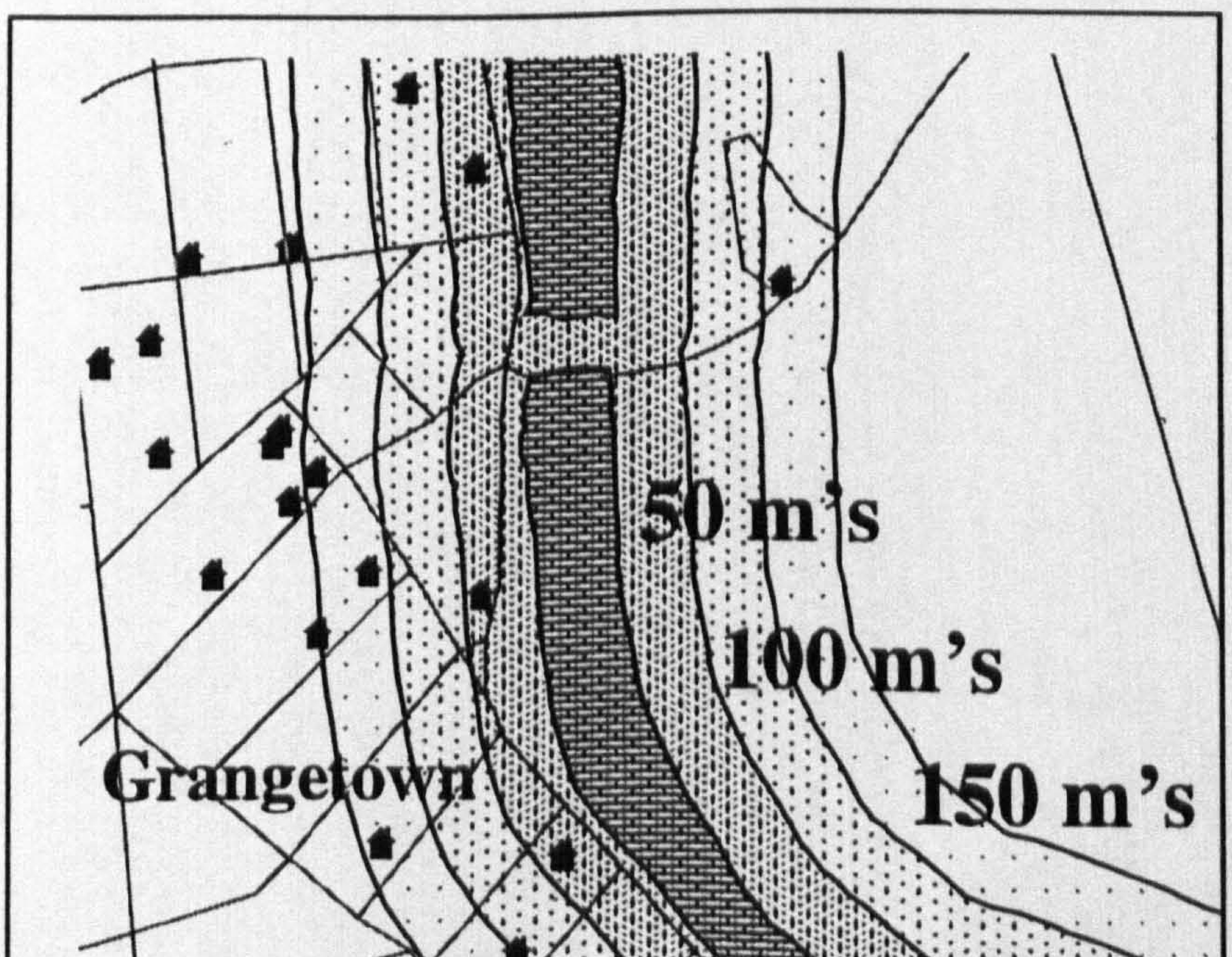
Table 6.6 summarises the non-residential land-uses hypothesized to affect property prices in the Inner Area of Cardiff. Their effects depend upon the distance from the property and their relative attractiveness as an amenity / disamenity. Both must be modelled simultaneously. This can be achieved in ARC / INFO by using the ACCESSIBILITY command, which calculates a measure of proximity as directly proportional to the supply of attributes at a

Figure 6.26

Distance Intervals from the Railway



Distance Intervals from the River Taff



given location, and inversely proportional to the distance away from it. This can be summarised by equation (6.1)

$$P_i = \sum_{i=1}^{n} W_j \, d_{ij}^{-\beta} \qquad 6.1$$

where: P_i is the proximity of property i
W_j is the attractiveness of externality j
d_{ij} is the distance between property i and externality j
β is the exponent for distance decay
n is the number of externalities being measured

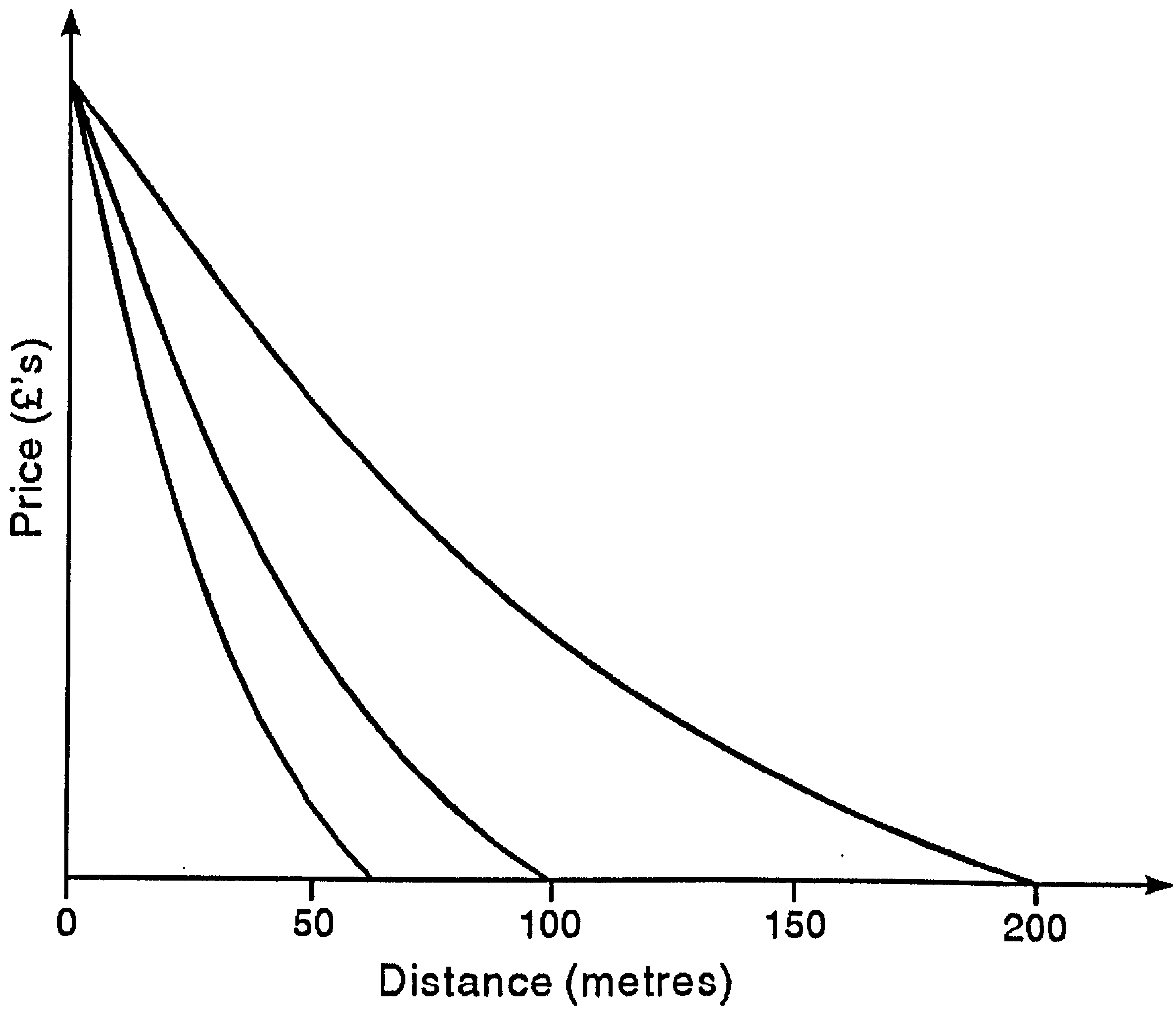
Hence, ACCESSIBILITY is a means of calculating externality effects, since it models both the magnitude of the effect and its proximity. The attributes of the externality are used to compute an attractiveness index, whilst the effects of distance are scaled using a distance decay function. Two distance decay functions are provided within ACCESSIBILITY: a power function that provides a gentle cut off to destinations and an exponential function that provides a steeper cut off. The latter is typically used for computing interactions over small distances, such as within a city. The process of finding the value of β in computing ACCESSIBILITY is called calibration, and is an important aspect of measuring the externality effect. If the exponent is small, the externality effect increases and vice versa. Since this value is not known *a priori*, different values were used to estimate different externality measures, and these are subsequently analysed in *Chapter Eight*.

The ACCESSIBILITY command was used to compute externality measures for the land-uses in Table 5.4, which can roughly be divided into parks, industrial areas and large institutional land-uses. The attractiveness index used in the computation was taken as the area of land squared. This was calculated within ARC / INFO using the AREA command. ACCESSIBILITY allows the computation of the effects of several externalities upon a single location simultaneously. Hence, separate measures for the effects of parks, light industrial, heavy industrial and institutional landuses were calculated.

The proximity measures of the remaining property level attributes in Table 5.1 were also modelled using ACCESSIBILITY, but with an attractiveness index set to unity. This means that the externality effect is solely determined by distance and not magnitude, since the construction of attractiveness indexes for features such as hospitals and local shops were considered too problematic and arbitrary. However, proximity measures to the River Taff

Figure 6.27

Three Hypothetical Street Externality Curves



and the railway lines were calculated in a slightly different way. By using previous studies as a yard stick (see *Chapter Three*), it was hypothesized that the externality effects associated with these features would be quite small. Four sets of buffer zones were generated at intervals of fifty metres from each feature - see Figure 6.26. The POINT-IN-POLYGON command was then used with the property coverage to determine which properties fell within fifty, one hundred, one hundred and fifty and two hundred metres of the river and railway lines respectively.

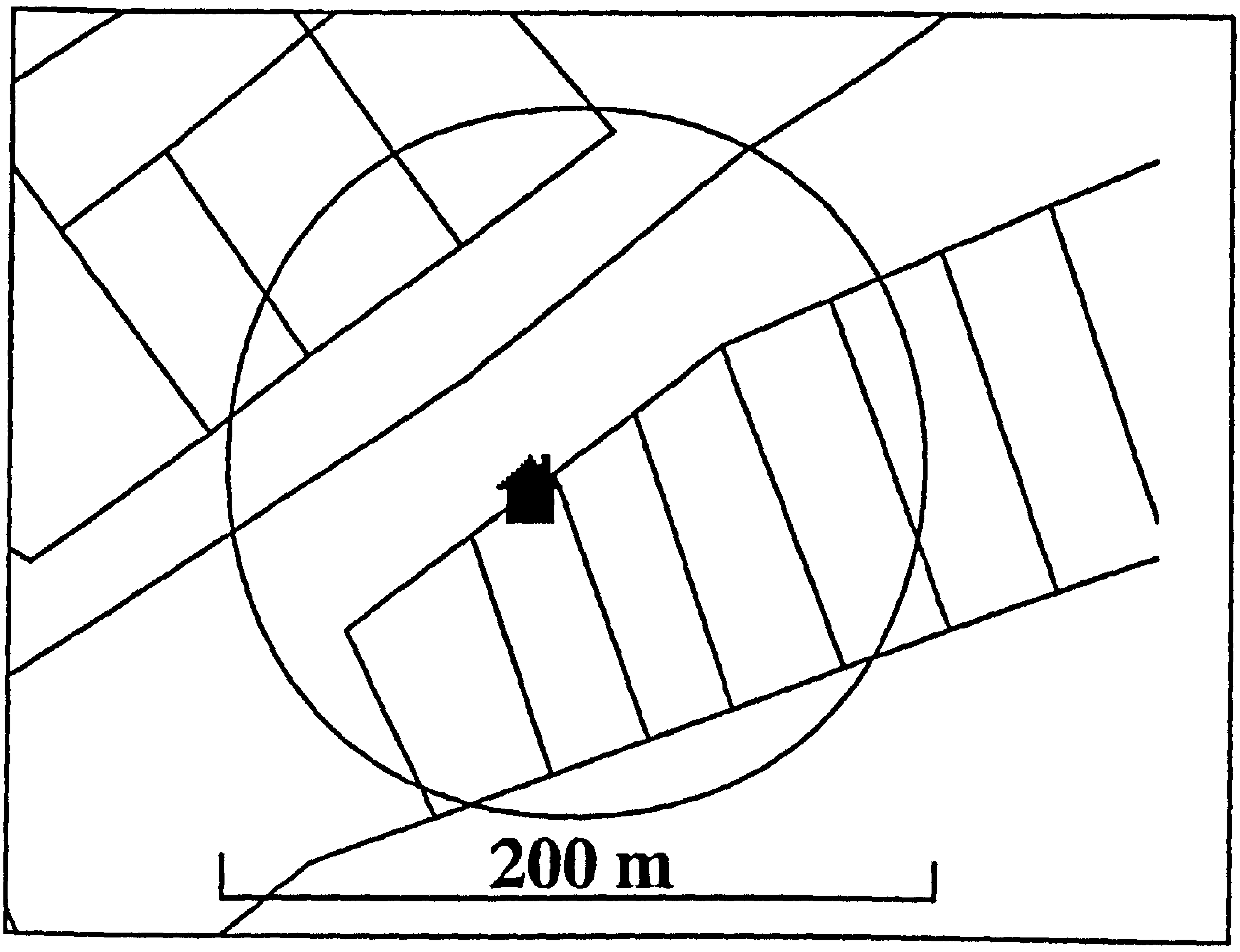
6.3.3.2 Street Level Externalities

Table 6.6 identifies the two sets of externalities that can be conceived to operate at street level. The first pertain to the environmental quality of individual streets, whilst the second relate to secondary school catchment areas. The generation of these externality effects are described below.

In *Chapter Five*, it was described how environmental quality data recorded in the CHCS was aggregated to sub-street level and then used to calculate two attributes of street environment: overall street quality and the specific impact of non-residential activity located in that street. It is anticipated that these two attributes will affect properties located within the immediate street. However, in *Chapter Three* it was argued that it is not known *a priori* how a locational externality diminishes with distance, and that the environmental quality of one street may also influence properties in adjacent streets, depending upon the perceptions of the buyer. Figure 6.27 is a graphical illustration of three hypothetical curves representing street environment externalities. The first curve is very steep, and the street environment has no influence upon properties beyond fifty metres of the externality. In this case, the street environment only affects properties in the immediate street. The second and third curve proposes that the externality has a gentler distance decay and thus influences properties in a wider area. Hence, three street quality externality effects were calculated for each property, using the two street quality attributes held in the street coverage. The first externality effect (the first curve) relates solely to the attributes held at the sub-street level, and the attributes were obtained by using the sub-street codes that links the sub-street data to the property coverage. The attributes that measure the remaining two street quality externality effects

Figure 6.28

Traditional Circular Buffer Zone



Topologically Sensitive Buffer Zone

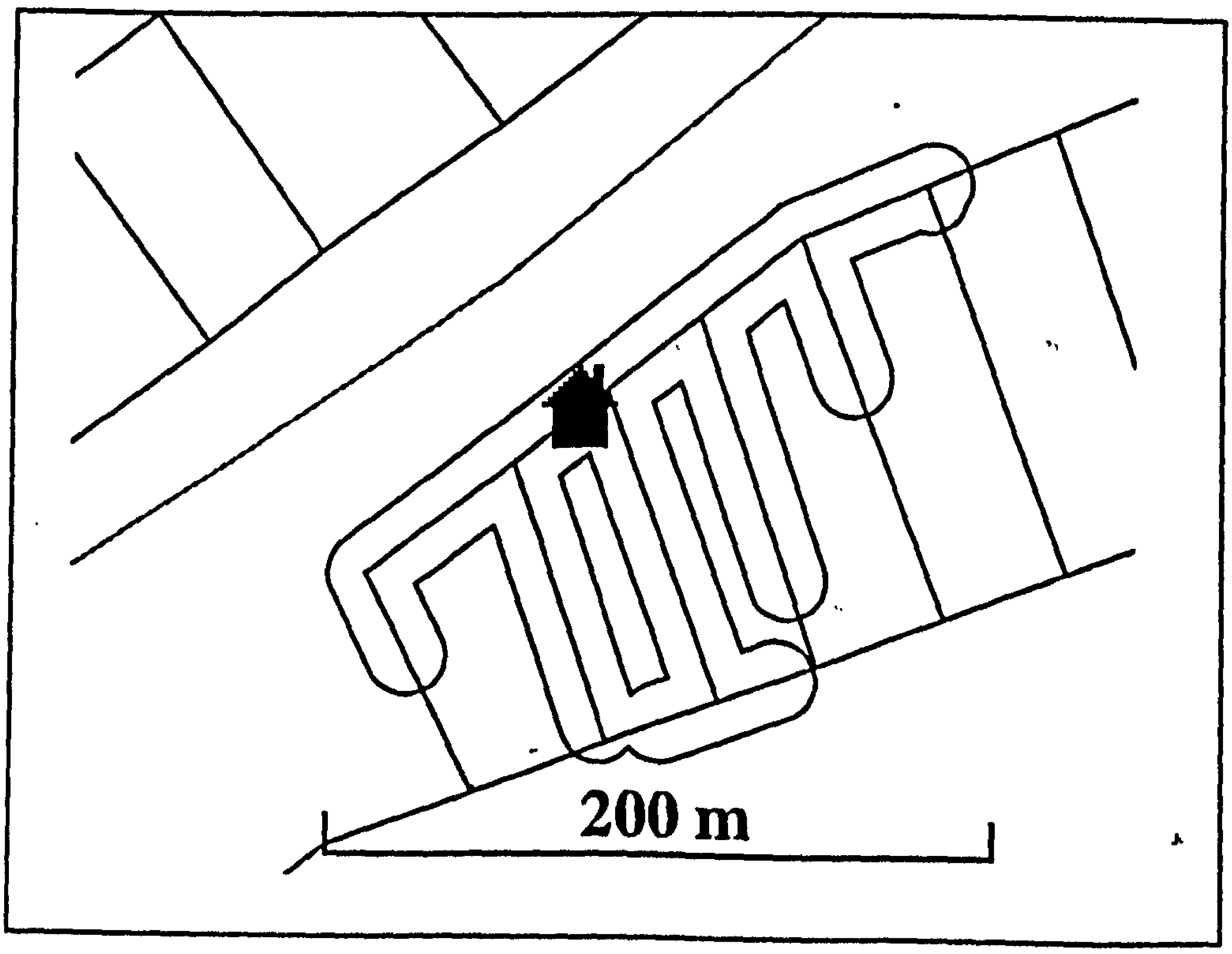
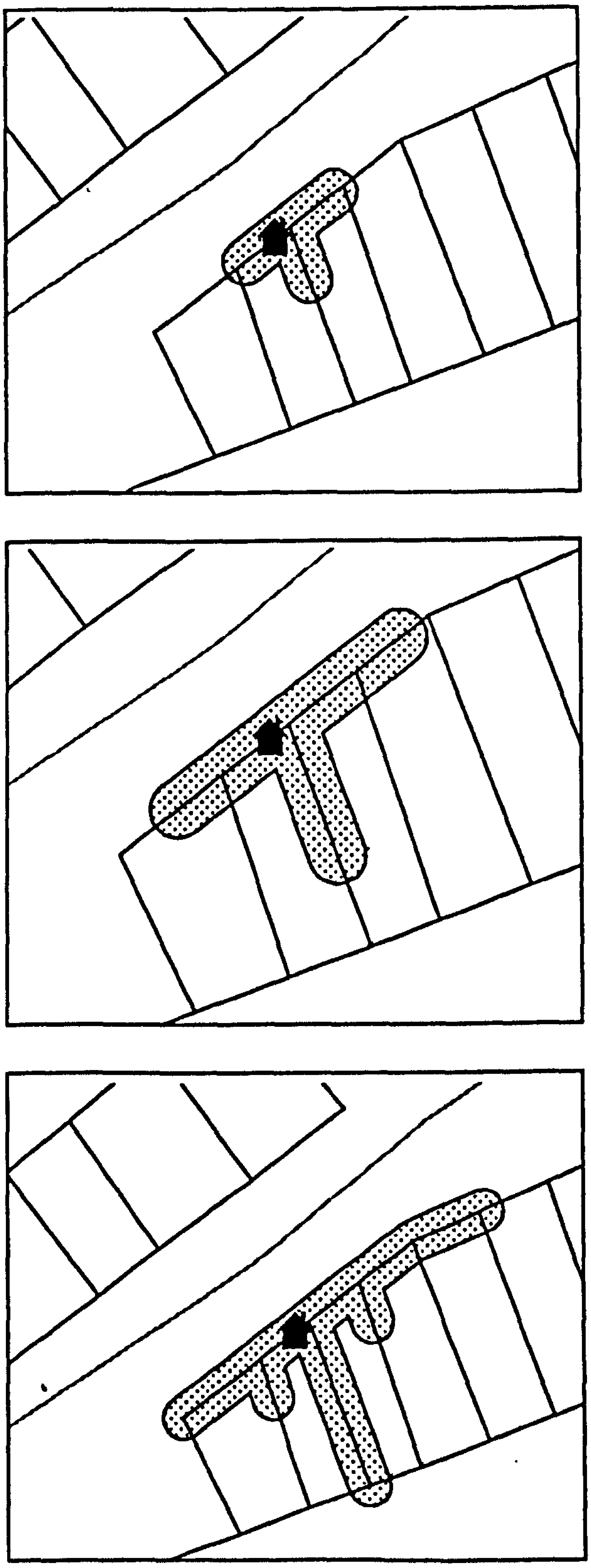


Figure 6.29

Street Externality Buffer Zones



50 metres

100 metres

200 metres

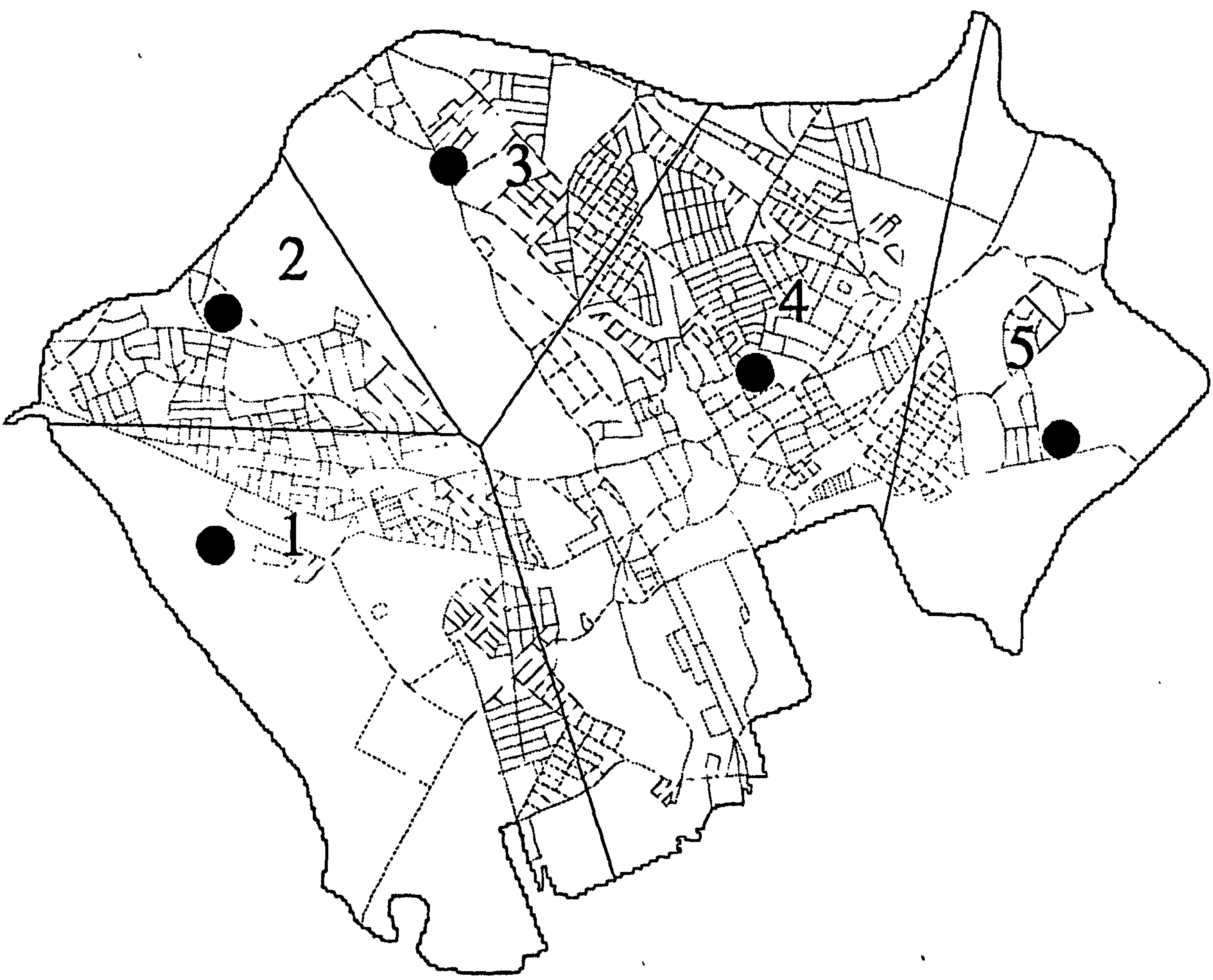
were extracted from the street coverage using buffer zones generated for each property in the property coverage.

Two methods were considered for generating these buffer zones. The first method used a simple circular buffer zone around a property. Unfortunately, this assumes that the urban area is continuous, and does not take into consideration physical boundaries such as railway lines, rivers, main roads and the orientation of the street network that may influence the buyers perception of the surrounding street environment. This is illustrated in the upper panel of Figure 6.28, in which a circular buffer cuts across a railway line and captures the attributes of the street environment on the opposite side, even though a buyer is likely to perceive these to be inconsequential and thus to have no influence upon property price.

Instead, a more sophisticated method of extracting the street quality data was used, which took into account the underlying topology of the urban area. This involved using the street network as the basis for the data extraction procedure. The ALLOCATE command in the NETWORK module allows routes to be calculated along the street network, using the length of the street as an impedance cost to limit its extent. These routes can then be used to generate buffer zones which lie exactly along the street network, and is illustrated in the lower panel of Figure 6.28. This ameliorates the problem of the buffer zone crossing boundaries, such as railways. These buffer zones can then be used to extract street quality data held in the street coverage. However, a problem arises when specifying the maximum extent of each buffer zone. Theoretically, this would depend upon the influence of street quality on the surrounding properties. Because the extent of this is not known *a priori*, it was decided to restrict the extent of each buffer to 100 metres and 200 metres respectively. These values represent the range of the majority of street lengths in the Inner Area. A further consideration was the influence of road junctions and turns in the road, since these could diminish the perception of the quality of adjacent streets. This can be accounted for in NETWORK by specifying an extra impedance cost for a turn in the network, which would limit the size of the buffer zone and thus the extent of the externality effect. Such an impedance cost would necessarily be arbitrary, and a value of 50 metres was chosen since this would have the effect of diminishing the perception of street quality by a third of the overall length. The result, illustrated in Figure 6.29, is a more realistic view to how street quality would be perceived from any particular house, as opposed to using circular buffers that does not take the typology into account. A POINT-IN-POLYGON analysis was then undertaken using each of the buffers to extract the two sets of attribute information

Figure 6.30

Inner Area Secondary School Catchments



1 Fitzalan High School
2 Cantonia High School
3 Cathays High School
4 St Teilo's High School
5 Willows High School

1 km

pertaining to street quality held at the sub-street level. Where necessary, the model value of these data were calculated for each attribute, and were subsequently attached to the property level point coverage. Hence each property had a measure of street quality based on four categories (poor, below average, above average, good) at three spatial resolutions (0 - 50m, 50 - 100m and 100 - 200m).

Secondary school catchment areas had to be approximated using the THIESSEN command. This converts the secondary school point coverage into a Thiessen polygon coverage, representing each individual schools catchment area. However, this will only be appropriate for those schools who base their intake upon the surrounding residential population. Table 5.6 contains information about each secondary school in the Inner Area, and this includes whether the school is Independent or under Local Authority control. Since it is typical for the former to be selective about their pupil intake, pupils need not live in the surrounding residential area to be allowed to attend. Hence, property prices are less likely to be influenced by these schools and so it would be inappropriate to construct Thiessen polygon coverages for these schools. Figure 6.30 illustrates the Thiessen polygon catchment areas for the remaining five secondary schools that serve the Inner Area, and each property was placed into a catchment area using the POINT-IN-POLYGON command.

6.3.3.3 HCS Area Level Externalities

Table 6.6 summarises a range of neighbourhood externalities. These externalities are basically blanket measures that will have an absolute effect upon all property prices within a HCS Area. The variables measuring the quality of local amenities had previous been attached to the HCS Area coverage in *Chapter Five*. HCS Area level landuse externalities were calculated within ARC / INFO using the INTERSECT command to calculate the proportion of non-residential landuses in each HCS Area. The operation computes the geometric intersection of the landuse coverages and the HCS Area coverage, and only preserves the areas common to both. This allowed the total area of open space and non-residential landuse in each HCS Area to be extracted, and the proportion that this represented to be calculated using the AREA command. Housing density was calculated by similar means using the ADDRESS-POINT property coverage as a means of determining the number of properties in a particular HCS Area. The final variable to operate at the HCS Area was the dummy variable measuring the percentage of Local Authority tenure. In the

Table 6.7

Inner Area Locational Attributes

Variable	Variable
Property Level	**Street Level (cont.)**
Accessibility to CBD	Street quality 0-50m: Above Average
Accessibility to M4 motorway	Street quality 0-50m: Good
Accessibility to railway stations	Street quality 50-100m: Poor
Proximity to hospitals	Street quality 50-100m: Below Average
Proximity to sports centres	Street quality 50-100m: Above Average
Proximity to community centres	Street quality 50-100m: Good
Proximity to institutional centres	Street quality 100-200m: Poor
Proximity to local shops	Street quality 100-200m: Below Average
Proximity to primary schools	Street quality 100-200m: Above Average
Proximity to secondary schools	Street quality 100-200m: Good
Proximity to Bute Park	Street non-residential landuse.
Proximity to parks / open space	Sch Catchment: Willows High School
Proximity to light industrial land-use	Sch Catchment: Fitzalan High School
Proximity to heavy industrial land-use	Sch Catchment: Cantonia High School
Rail 0 -50m	Sch Catchment: Cathays High School
Rail 50 - 100m	Sch Catchment: St Teilo's High School
Rail 100 - 150m	**HCS Area Level**
Rail 150 - 200m	Percentage Local Authority tenure
River 0 - 50m	Percentage of open space
River 50 - 100m	Percentage of non-residential land-use
River 100 - 150m	Housing density
River 150 - 200m	Quality of local shops
Street Level	Quality of local public transport
Road Type: Primary	Quality of local sport facilities
Road Type: Secondary	Quality of local parks
Road Type: Residential	Quality of local community facilities
Road Type: Cul-de-sac / Close	**Neighbourhood Level**
Street quality 0-50m: Poor	Social economic class
Street quality 0-50m: Below Average	

Cardiff housing market study, this operates at the ED level. These data were aggregated to HCS Area level using the POINT-IN-POLYGON technique described in *Chapter Five.*

6.3.3.4 Community Level Externalities

The social economic class variable computed using principal components analysis was used as a measure of social composition of each community, whilst prestige and desirability were captured using the community boundaries. Table 6.7 is a summary of all the locational variables constructed for the Inner Area study. This illustrates the importance of the GIS in generating locational externality measures, and indicates a possible reason why the previous studies that have not a GIS have failed to capture the complexity of locational attributes.

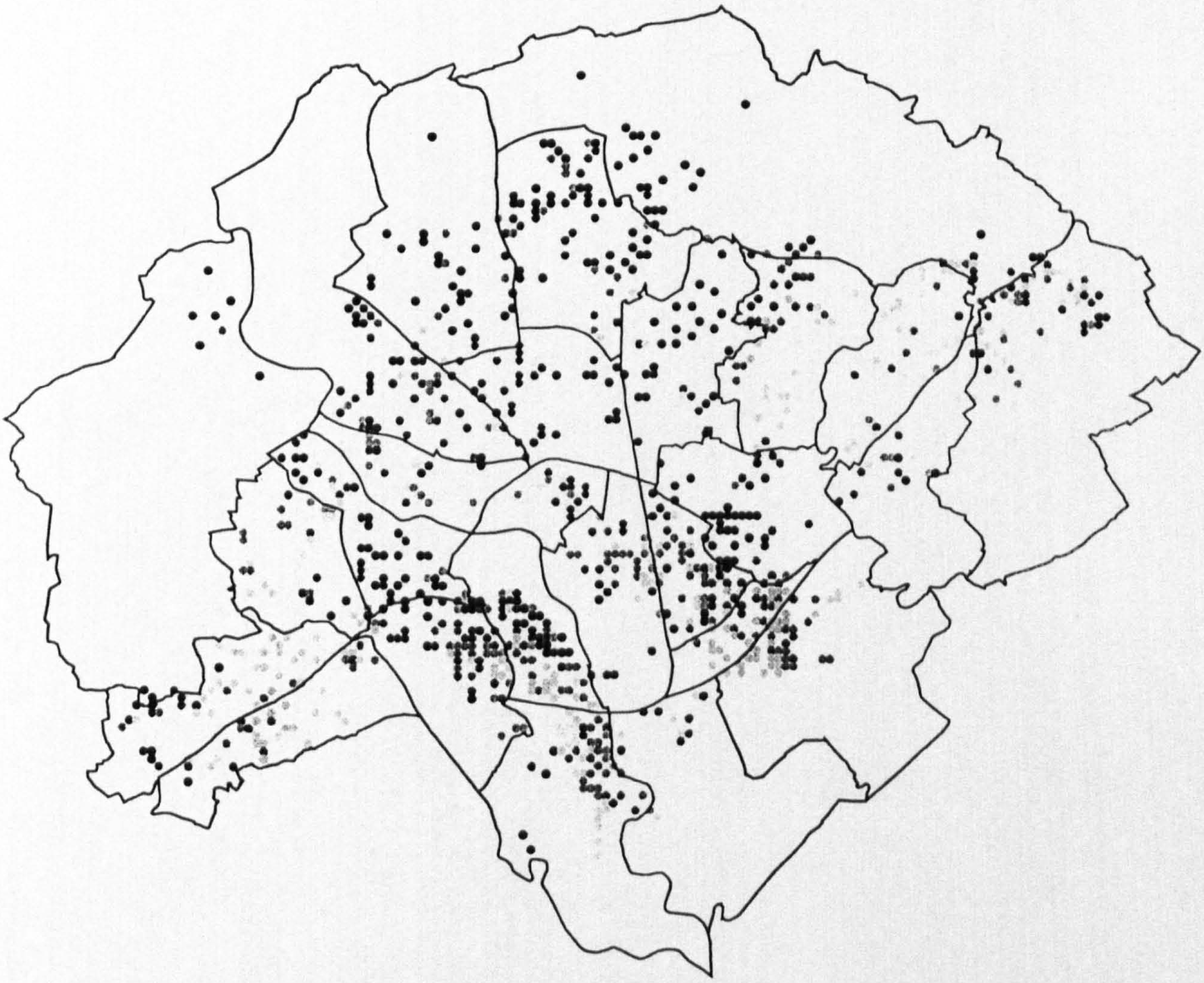
Section 6.4 The Geography of Housing Attributes

The previous sections have described the housing attribute data in detail. This section continues this theme by examining the geographical distribution of several of these attributes within Cardiff. In particular, geographical variations in house price, dwelling type and property size will be investigated, since these will have an important contribution in the interpretation of the subsequent hedonic models.

Figure 6.31 illustrates the geographical variation of house prices in Cardiff in February 1995, taken from the house price survey. These have been banded into quartiles, and it is evident from their distribution that house prices display a marked clustering within communities. The most expensive twenty five percent of properties in the sample are predominantly located in the communities to the north of the city. In particular, Lisvane and St Mellons, Llanishen, Rhiwbina, Cyncoed and Heath all contain a significant proportion of expensive properties. Closer to the city centre, Llandaff, Roath and Riverside are also notable. The cheapest twenty five percent of properties tend to be located within the more peripheral communities, such as Ely and Caerau in the west, and Trowbridge, Llanrumney and Pentwyn in the east. More average priced properties are generally found in the Inner Area communities, but also in some suburban locations such as Fairwater and Rumney. However, this is a generalisation, and the map demonstrates clearly that nearly all the communities contain properties from each price band. This is confirmed in Table 6.8, which

Figure 6.31

The Geography of Cardiff House Prices - February 1995



Quartiles (Pounds)

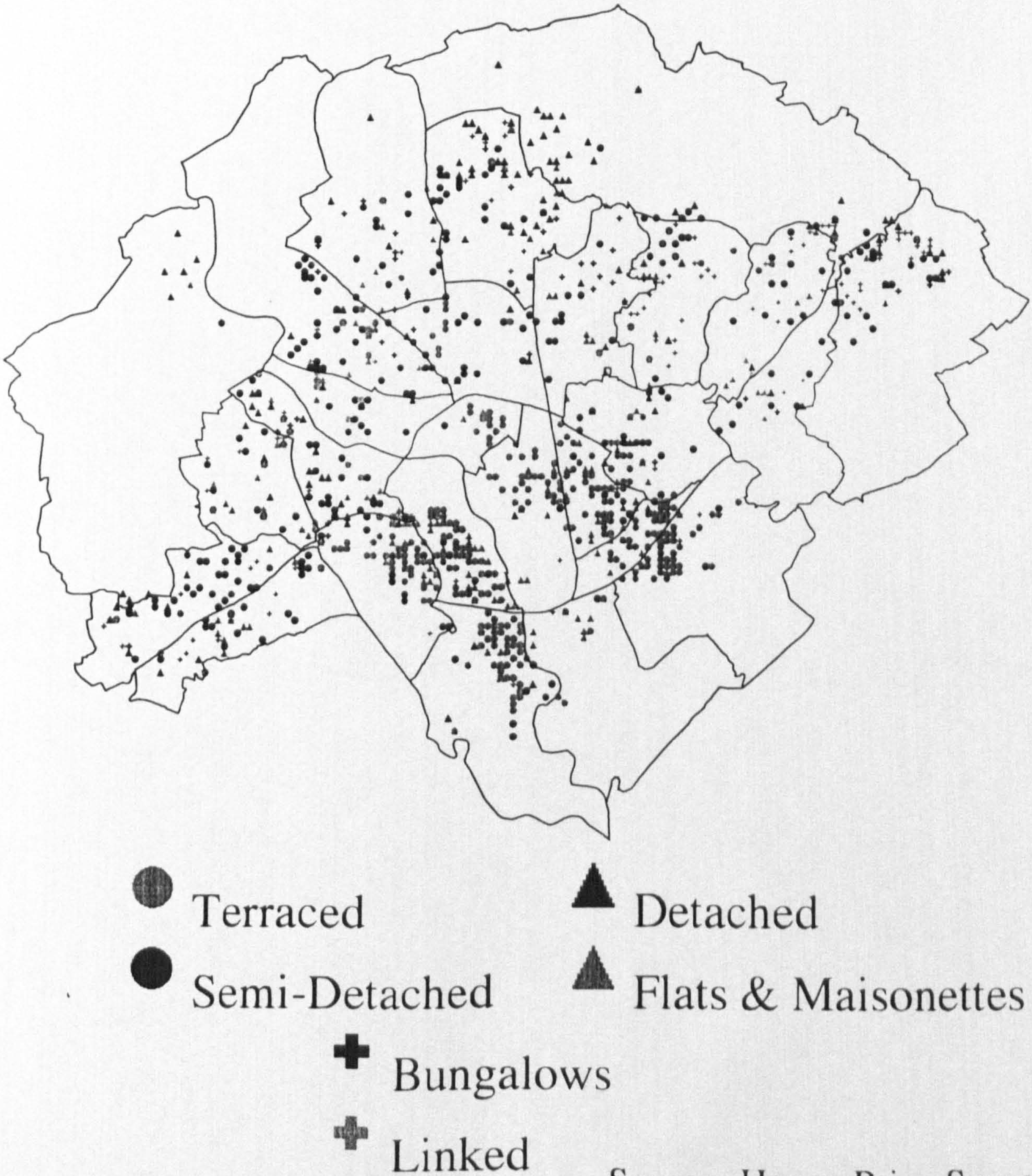
- 15950 - 39950
- 39951 - 52950
- 52951 - 75950
- 75951 - 330000

1 km

Source: House Price Survey

Figure 6.32

The Geography of Dwelling Types in Cardiff



Source: House Price Survey

summaries the range of prices within each community. The greatest price range occurs in the large, suburban communities where the housing stock is more diverse, such as Lisvane and St. Mellons, as opposed to the smaller, more homogeneous inner-city communities, such as Gabalfa and Splott. A better picture of house price variation can be revealed by examining the interquartile ranges, since these represent the range of prices that are more typical within each community. For just under a third of the communities, this price range also represents the average price for Cardiff (£ 65,888), whilst for a half, this price range falls below the city average (highlighted in italics). The remaining communities have a typical price range above the Cardiff average (highlighted in bold), and with the exception of Radyr and St Fagans, form a contiguous belt to the immediate north of the Inner Area.

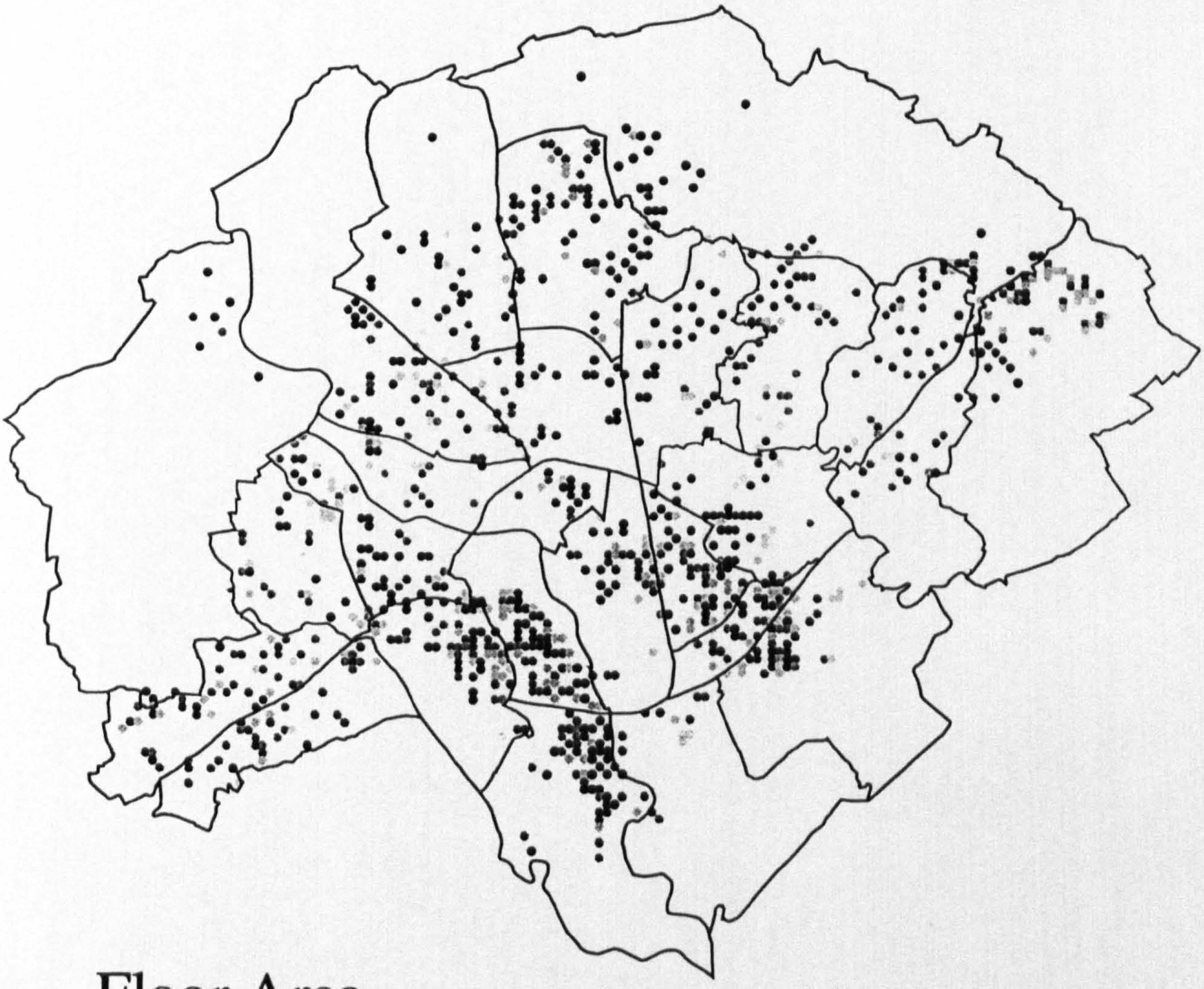
Figure 6.32 depicts the geographical variation in dwelling type. Similar to house price variation, dwelling types are also clustered within certain communities, although this clustering is naturally more concentrated. For instance, terraced properties, flats and maisonettes are typically located with the Inner Area communities, with semi-detached, detached and linked properties located within more suburban locations, representing the historical growth of the city. In more detail, certain dwelling types appear to be restricted to certain communities. Detached houses tend to be located within Llandaff, Rhiwbina and Lisvane and St. Mellons, bungalows in Cyncoed, and linked properties in the peripheral estates of Pentwyn, Trowbridge, Ely and Caerau. The only ubiquitous dwelling type would appear to be semi-detached housing, although this is still under-represented in the Inner Area as would be expected.

The geographical variation in property size is described in Figure 6.33 and Table 6.8. This shows a very similar geography to that of house prices, as might be expected, although there are some interesting departures. Firstly, a significant proportion of properties in the Inner Area communities fall within the upper quartile of property size. This is especially evident in Grangetown and Adamsdown. However, house prices in these communities tends to be concentrated in the lower quartiles. Those communities in which the mean property size falls below the Cardiff average tend to be concentrated in the peripheral estates such as Ely and Caerau, and Rumney and Trowbridge and the smaller Inner Area communities such as Adamsdown and Butetown.

Figure 6.34 summaries the Table 6.8 by depicting the geography of mean house prices within Cardiff, and also the typical property. Again, this illustrates the north / south

Figure 6.33

The Geographical Distribution of Dwelling Size



Floor Area
Quartiles (sq-ft)

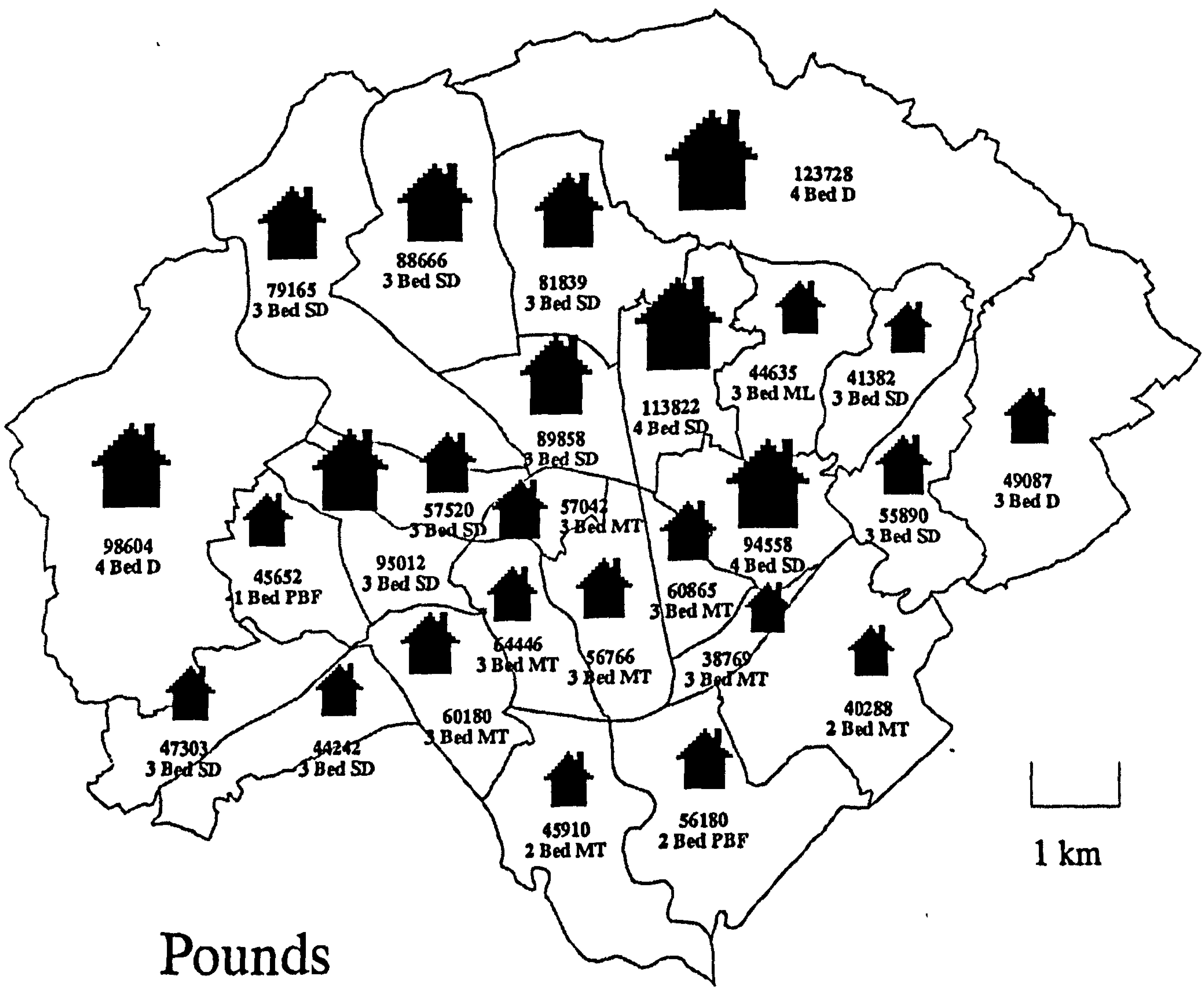
- 0 - 529.3
- 529.3 - 685.2
- 685.2 - 860.8
- 860.8 - 2843.96

1 km

Source: House Price Survey

Figure 6.34

Mean House Price and Typical Dwelling Type by Community



Pounds

40000

80000

120000

MT Mid-Terrace

SD Semi-Detached

PBF Purpose Built Flat

D Detached

Source: House Price Survey

differentials in house prices, with the most expensive communities forming a contiguous band from the north of the Inner Area, to the edge of the city. The peripheral communities to the south and the west of the city contain the cheapest properties. The predominance of three bedroomed terraced properties is evident as the typical in the Inner Area, whilst the ubiquitous three bedroomed semi-detached house is the typical property for the remaining communities. The exceptions are Lisvane and St. Mellons and Radyr and St. Fagans where the four bedroomed detached is the typical housing stock, and Fairwater and Butetown where one and two bedroomed flats predominate.

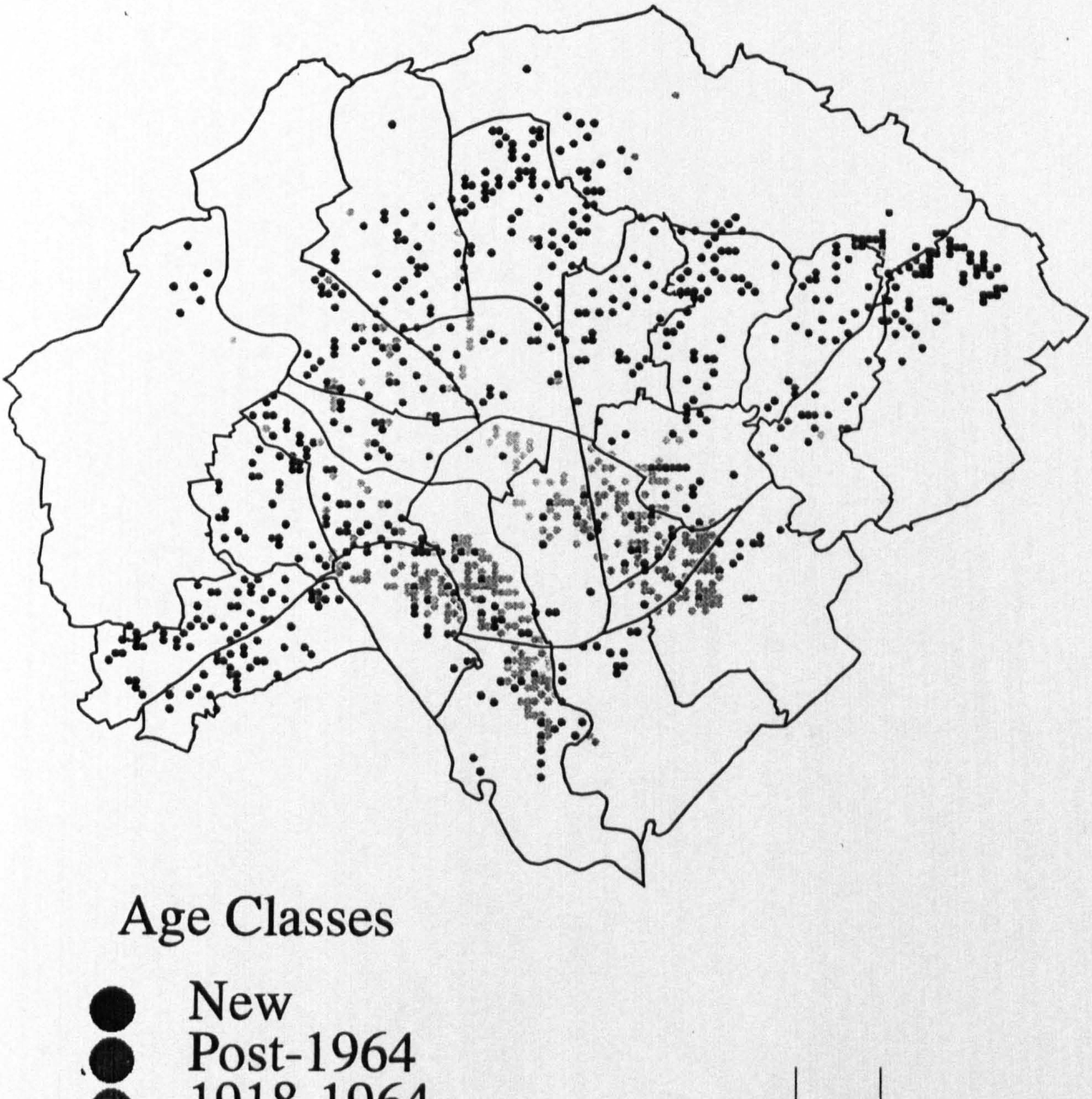
Table 6.8

A Summary of House Price and Floor Area Variation by Community

Community	House Prices					Floor	Modal
	Min	Max	Inter	quartile	Mean	(Sq-Ft)	Property
Adamsdown	26950	125000	*31350*	*41950*	38769	679.40	3 bed MT
Butetown	29950	89500	40475	71450	56180	594.80	2 bed PBF
Caerau	23500	120000	*34475*	*49450*	44242	646.45	3 bed SD
Canton	19950	159950	*47950*	*62838*	60180	745.94	3 bed MT
Cathays	19950	225000	*37950*	*63950*	56766	609.48	3 bed MT
Cyncoed	28500	280000	**75000**	**139950**	113822	897.44	4 bed SD
Ely	22950	112500	*32950*	*57487*	47303	665.72	3 bed SD
Fairwater	17500	115000	*34950*	*48950*	45652	563.15	1 bed PBF
Gabalfa	34950	69950	*50700*	*65713*	57042	793.57	3 bed MT
Grangetown	22500	115000	*35950*	*51238*	45910	742.62	2 bed MT
Heath	44950	235000	**67500**	**84988**	89858	828.94	3 bed SD
Landaff	24950	295000	52738	129950	95012	999.68	3 bed SD
Lisvane and St Mellons	32950	330000	55950	159972	123728	838.91	4 bed D
Llandaff North	33250	104950	*46975*	*62000*	57520	657.55	3 bed SD
Llanishen	32550	229000	49950	109750	81839	699.51	3 bed SD
LLanrumney	19950	89950	*33950*	*42950*	41382	650.72	3 bed SD
Pentwyn	15950	84950	*32500*	*55950*	44635	617.87	3 bed ML
Plasnewydd	18000	385000	41000	77362	60865	763.82	3 bed MT
Radyr and St Fagans	45000	145000	**76375**	**119950**	98604	1007.50	4 bed D
Rhiwbina	27950	245000	62450	98125	88666	814.07	3 bed SD
Riverside	21950	255000	40700	75238	64446	799.83	3 bed MT
Roath	41950	295000	**66988**	**118125**	94558	972.93	4 bed SD
Rumney	24950	89995	40450	69475	55890	694.42	3 bed SD
Splott	24500	59500	*35950*	*43950*	40288	621.83	2 bed MT
Trowbridge	21950	89950	*37950*	*57950*	49087	596.57	3 bed D
Whitchurch and Tongwylais	34995	228950	49950	86450	79165	784.48	3 bed SD
Cardiff	15950	330000	39950	75963	65888	739.51	3 bed MT

Figure 6.35

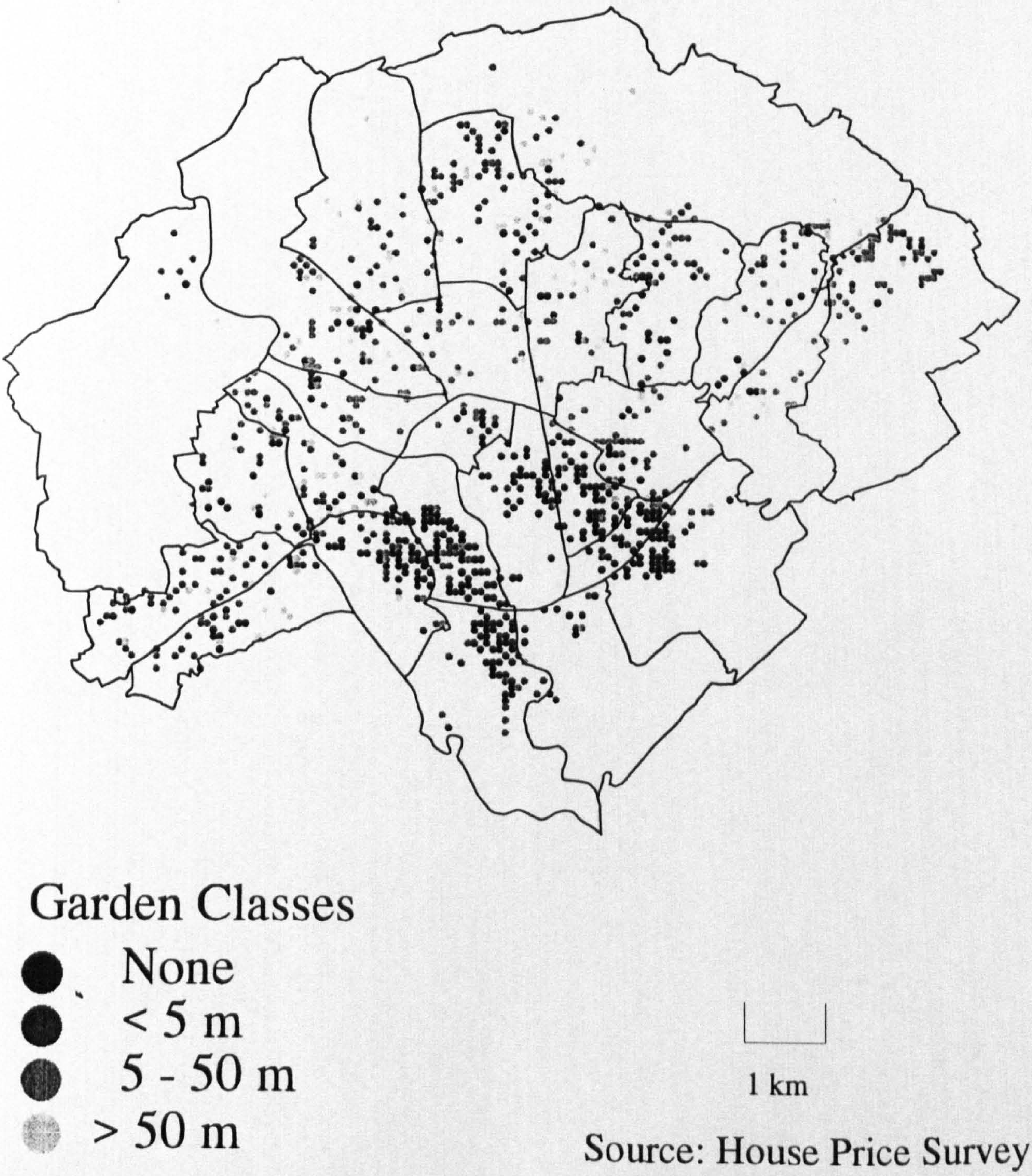
The Geographical Distribution of Dwelling Age



Source: House Price Survey

Figure 6.36

The Geography of Garden Size



Source: House Price Survey

Figure 6.37

Variation in Age of Properties with Distance from Cardiff City Centre

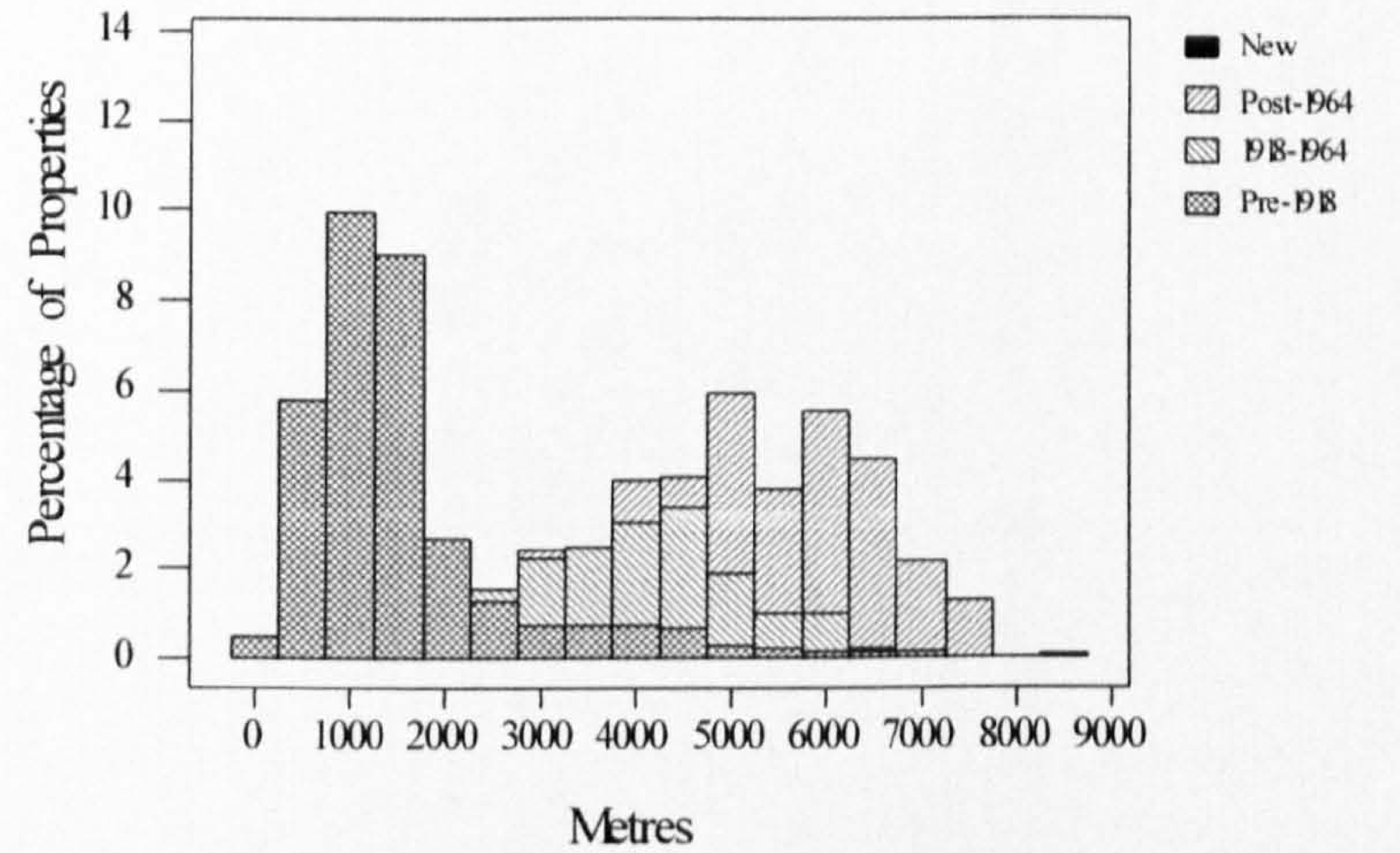


Figure 6.38

Variation in Garden Size with Distance from Cardiff City Centre

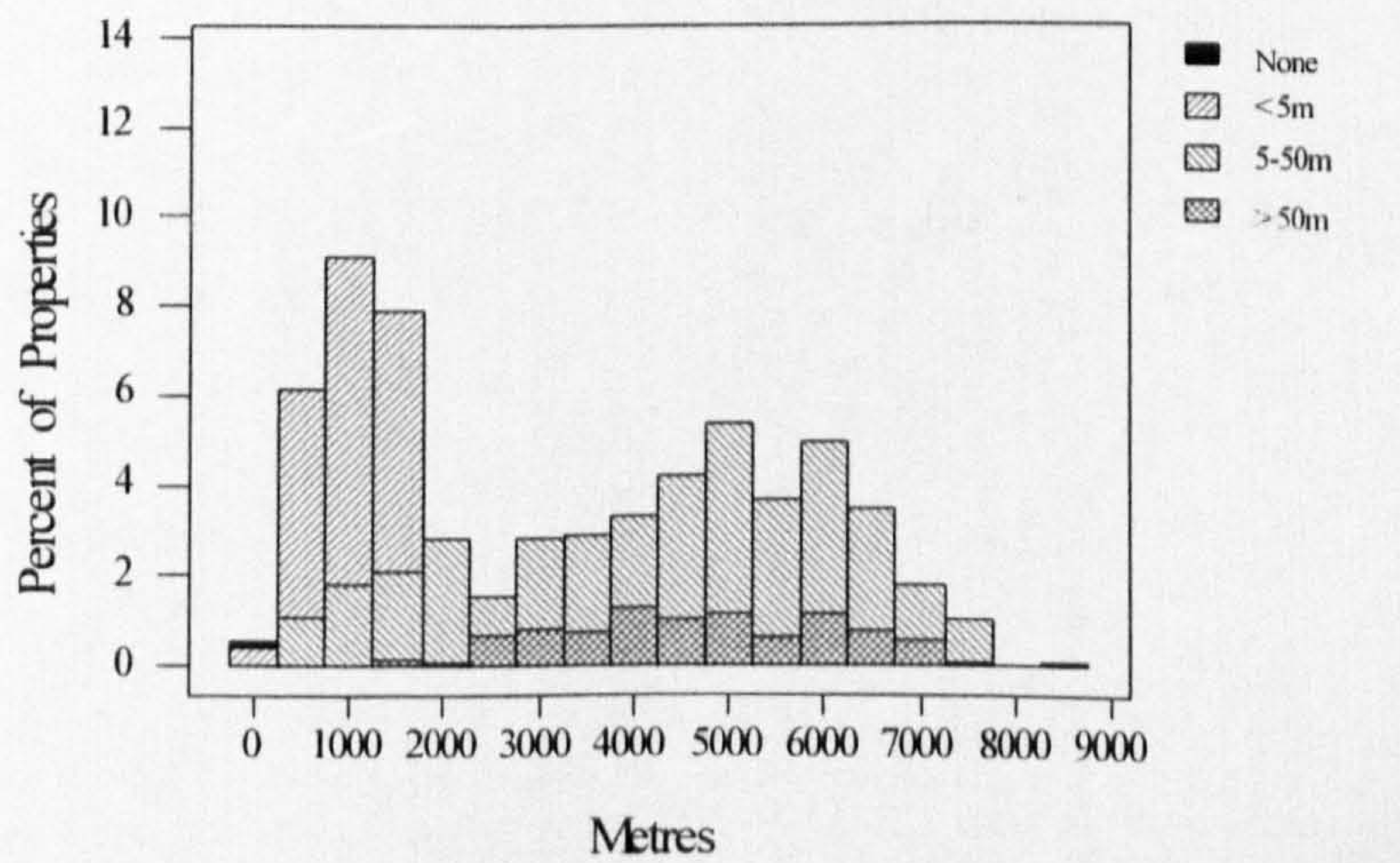


Figure 6.34 summaries the above by depicting the geography of mean house prices within Cardiff, and also the typical property. Again, this illustrates the north / south differentials in house prices, with the most expensive communities forming a contiguous band from the north of the Inner Area, to the edge of the city. The peripheral communities to the south and the west of the city contain the cheapest properties. The predominance of three bedroomed terraced properties is evident as the typical in the Inner Area, whilst the ubiquitous three bedroomed semi-detached house is the typical property for the remaining communities. The exceptions are Lisvane and St. Mellons and Radyr and St. Fagans where the four bedroomed detached is the typical housing stock, and Fairwater and Butetown where one and two bedroomed flats predominate.

The geography of two other structural attributes that are worth consideration are age and garden size. Since these attributes are bound up with the historical development of the city, their distribution may display definite spatial patterning, the conjecture being that they may be a source of spatial multicollinearity in the subsequent models. Figure 6.35 reveals that the age categories form distinctive bands out from the city centre, with only slight evidence of modern redevelopment in the Inner Area. However, Figure 6.36 suggests that this spatial patterning is less obvious for garden size, although the majority of properties in the Inner Area have a garden less than five metres in length reflecting the prominence of terraced housing. The relationship of age and garden size with distance from the city centre is established more formally in Figures 6.37 and 6.38. The former confirms the impression given by Figure 6.35 of distinctive age bands, with only slight evidence of age variations in the communities mid-distance from the city centre and the suburban fringe, whilst the latter suggests a more varied distribution of garden sizes throughout the city. The implications of the spatial relationship between age and distance are considered in more detail in the next section.

Section 6.5 Preliminary Hedonic Models

6.5.1 Introduction

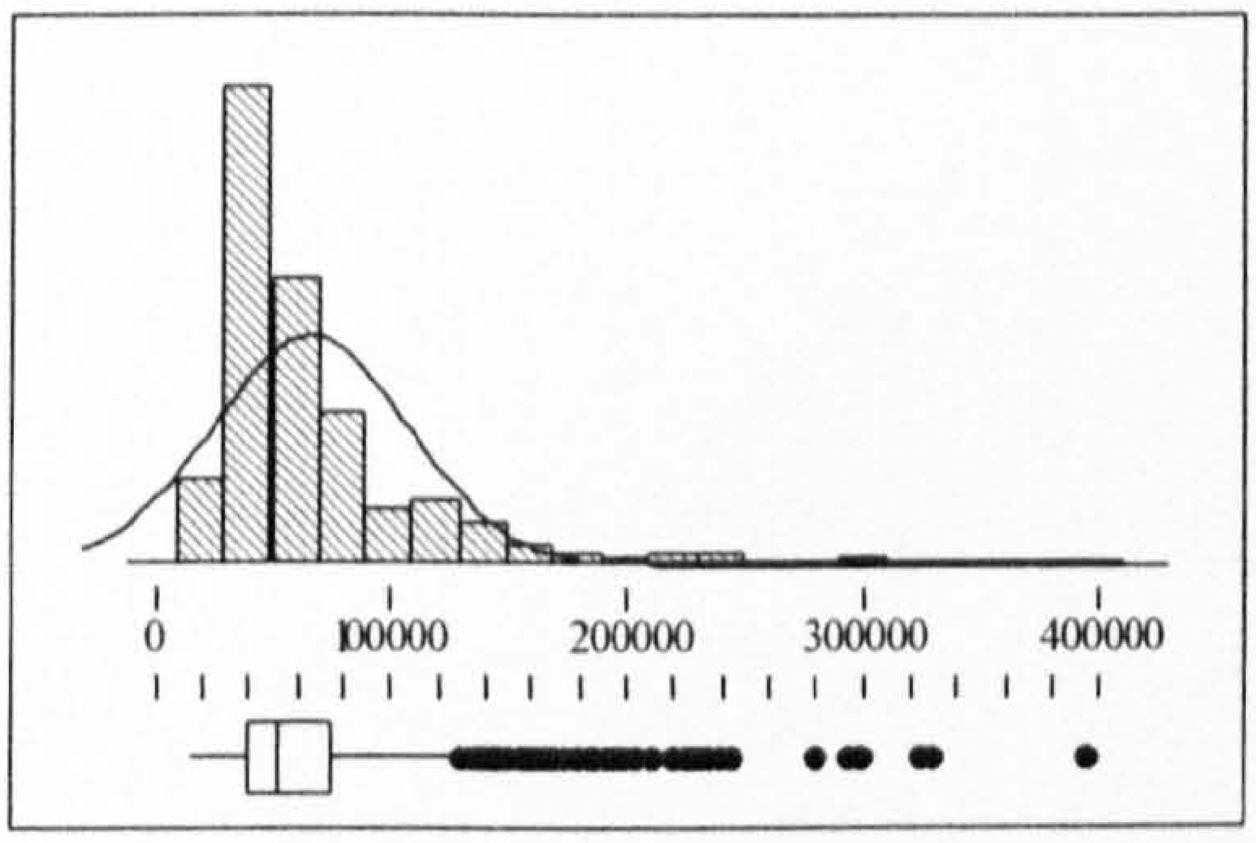
This chapter has so far been concerned with generating and describing the housing attribute data. *Chapter Seven* and *Chapter Eight* will discuss how these data were used to estimate hedonic models for the Cardiff housing market and the Inner Area respectively. However, prior to this, several preliminary hedonic models were estimated in an investigative capacity, as a means of exploring how the variables enter the model, and how the model explains the underlying structure of the data. This is an important part of the model building process, and one that is often neglected. It follows an iterative process that involves fitting a series of models to the data, checking that the assumptions are not violated, and refitting the models if necessary. The regression line of the final model should explain all the structure in the data. Hence, the aim of this section is not to explore the implications of the hedonic models in detail - this is investigated in the subsequent chapters - but to move towards some basic hedonic model which uncovers the key features of the data, whilst not violating the statistical assumptions behind the technique. This is called 'building a regression model' (Dunn, 1989).

At this preliminary stage of the investigation, the traditional specification of hedonic model (see *Chapter Two*) was estimated for the whole of the Cardiff housing market:

$$P_i = \alpha X_i + \Sigma \beta_k S_{ki} + \Sigma\gamma_q L_{qi} + \varepsilon_i X_i \qquad \textbf{6.2}$$

Where:
$i = 1, ..., N$ is the subscript denoting each property;
P_i is the price of property i;
$k = 1, ..., K$ is the number of structural attributes;
$q = 1, ..., Q$ is the number of locational attributes;
α, β, γ and ε are the corresponding parameters;
X_i is a column vector which consists entirely of ones.

This specification has several advantages when beginning to build an hedonic model. Firstly, it is simple and easy to estimate, but, as was demonstrated in *Chapter Two*, it forms the basic model upon which more complex hedonic models are built. Secondly, it employs all the structural variables, and since these are common to all the subsequent hedonic models in both *Chapters Seven* and *Eight*, it would appear sensible to understand in detail how they enter the model. More specifically, it can be hypothesized that house size will be the most important factor in determining house price, and this may enter the model in two distinct ways. Firstly, total floor area may be significant, and be modelled accordingly. Alternatively, average floor area may be more significant in determining price, and this can be regarded in terms of bedroom, recreation room and kitchen floor area. This suggests that



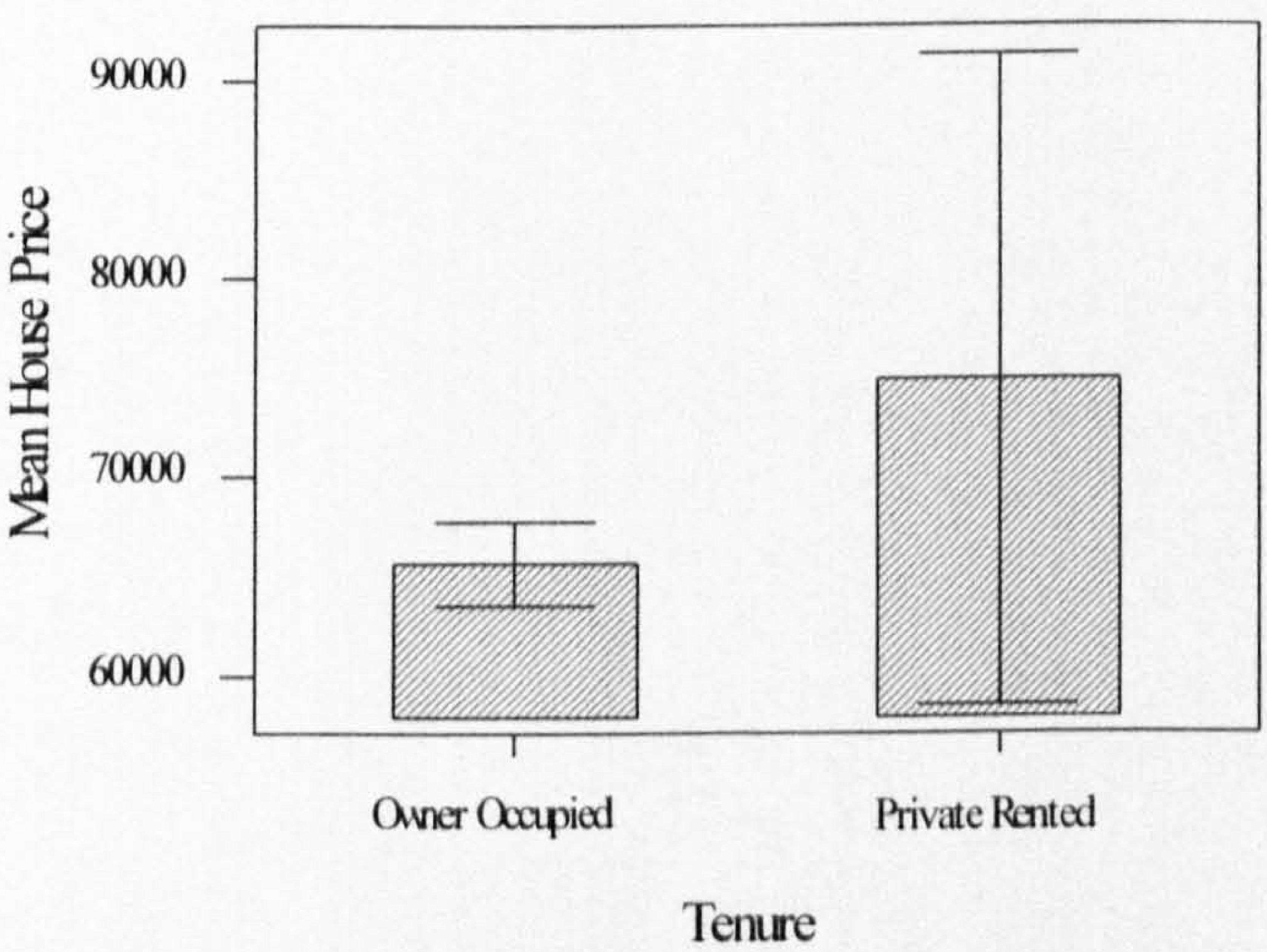
Mean	65805
Std Dev	41214
Skewness	3
Kurtosis	9
n of obs	1478
Min	15950
1st Quartile	39950
Median	52950
3rd Quartile	75950
Max	395000

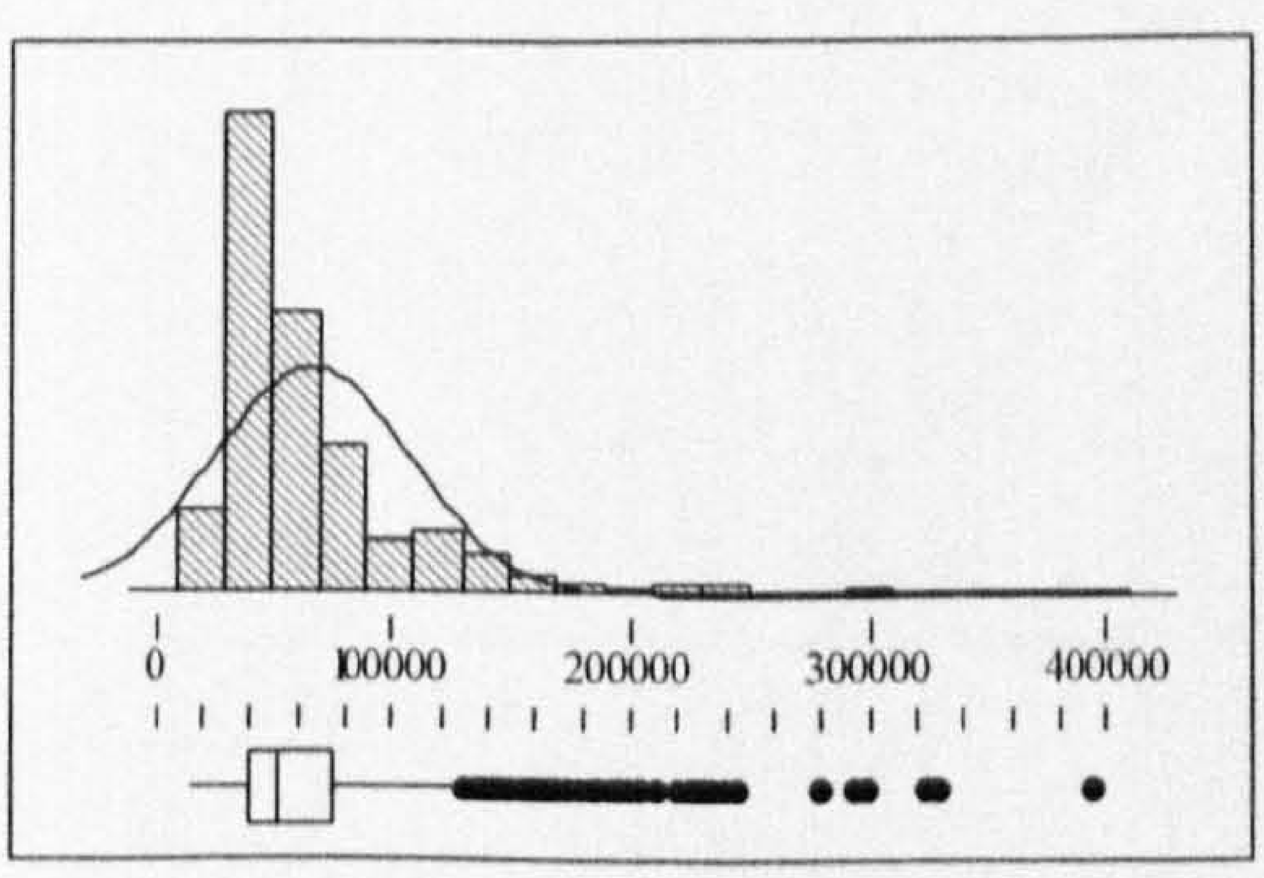
Figure 6.39

House Price Distribution for Cardiff

Figure 6.40

ANOVA of Owner Occupied and Private Rented Property Prices





Mean	65644
Std Dev	41146
Skewness	3
Kurtosis	9
n of obs	1447
Min	15950
1st Quartile	39950
Median	52000
3rd Quartile	75950
Max	395000

Figure 6.41

House Price Distribution for Owner Occupied Sector two distinct hedonic models should be explored in the preliminary model building stage, before more complex spatial models are built.

A disadvantage of the traditional specification is that it applies to the whole housing market, and therefore assumes that no sub-markets exist and hence that no spatial drift occurs. If this assumption is unrealistic, then heteroscedasticity will be expected in the model. Moreover, since this preliminary analysis investigates the dataset for the whole of Cardiff, the locational attribute variables used are restricted to those constructed using principal components analysis and the simple measure of accessibility to the city centre. This implies that the preliminary hedonic models may also suffer from omitted variable bias. However, since the primary aim is two correct for violations of the first three assumptions, these potential problems may not be too important.

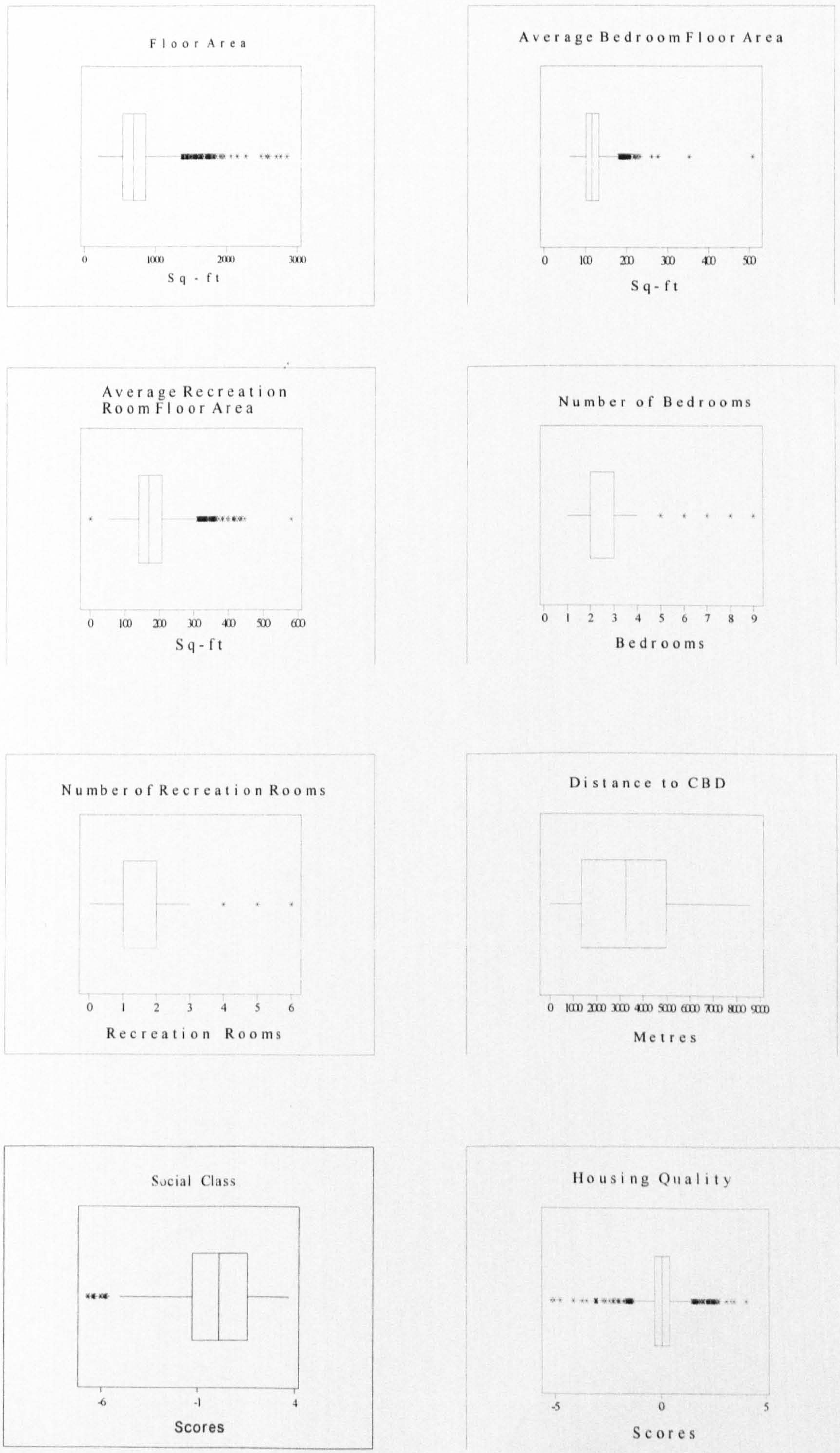
6.5.2 Exploratory Data Analysis

Prior to this exploratory model building exercise, it is beneficial to understand the characteristics of the data in more detail, especially with respect to the underlying distributions. As has been demonstrated by Anscombe (1973), an adequate descriptive summary of a data set is necessary if the results of statistical analysis are to be interpreted and understood correctly. In particular, it is essential that any problems incurred by potentially anomalous data are addressed. Since this preliminary study concerns data pertaining to the whole of the Cardiff housing market, the structural variables in Table 6.1 and the locational variables constructed from the census data.

6.5.2.1 The Dependent Variable

Examination of the dependent variable is very important, since the discovery of an unusual distribution or outliers may be a hint of likely problems in model building. It is also important to check for any major departures from normality. Figure 6.39 shows that half of the houses fall into the range of £39950 - £75950, whilst the majority are below £100 000. The box plot indicates that the distribution is highly skewed towards higher priced properties, with several in excess of £150 000 - more than twice the mean. The kurtosis indicates that this has resulted in making the plot flatter than normal. Although the size of the sample (c. 1500 properties) means that this departure from normality should have negligible effect upon the OLS regression, the estimates may be unduly influenced by the

Figure 6.42 Box Plots of the Continuously Distributed Variables



higher priced properties. The sample also contains sales data relating to thirty 'investment properties' - properties that are in the private rented sector, and are being sold on by the landlord. Since these properties are generally bought with view for continued renting, it can be expected that they operate under different market conditions than owner occupied properties, and hence will be priced differently. Although a simple one-way analysis of variance failed to prove that the means were significantly different at a 5% level (F-statistic = 1.50, critical value $F_{1,1476} = 3.84$ at 5% significance level), Figure 6.40 indicates that the variance of the investment properties is significantly bigger than those in the owner occupied sector. This suggests that the inclusion of investment properties within the sample may cause heteroscedasticity, which may be the case if supply and demand mechanisms for each sector are different. Hence, investment properties were omitted from the sample. Figure 6.41 describes the distribution of house price without the investment properties. The effect of omitting the investment properties has been to lower the mean and median house price, although the distribution remains positively skewed. This may indicate that the dependent variable would benefit from a transformation to remove this skewness. This is a common procedure in previous studies and is discussed in more detail in the next section.

6.5.2.2 Independent Variables

I. Box Plots

These are an effective means of graphically summarising continuously distributed data, particularly with respect to data outside the interquartile range. Figure 6.42 are box plots of the continuously distributed variables. Most of the variables are positively skewed, with a significant number of very large properties, suggesting possible outliers in the subsequent hedonic analysis. The interquartile ranges are quite small in comparison, suggesting that most properties are similar which can probably be attributed to the large number of terraced properties in the sample. The interquartile range of average bedroom size is particularly compact, supporting the findings of Figure 6.5. The number of bedrooms and recreation rooms appear to be roughly normally distributed, with only a few outlying observations. An interesting plot is that of housing quality. The distribution is negatively skewed, with most the observations either tightly clustered around the mean, or on the extremes of the distribution. Overall, the plots suggest that the majority of the independent variables are roughly normally distributed, although there are several potentially anomalous observations that may pose problems in the subsequent hedonic model.

Figure 6.43

House Price Against Floor Area

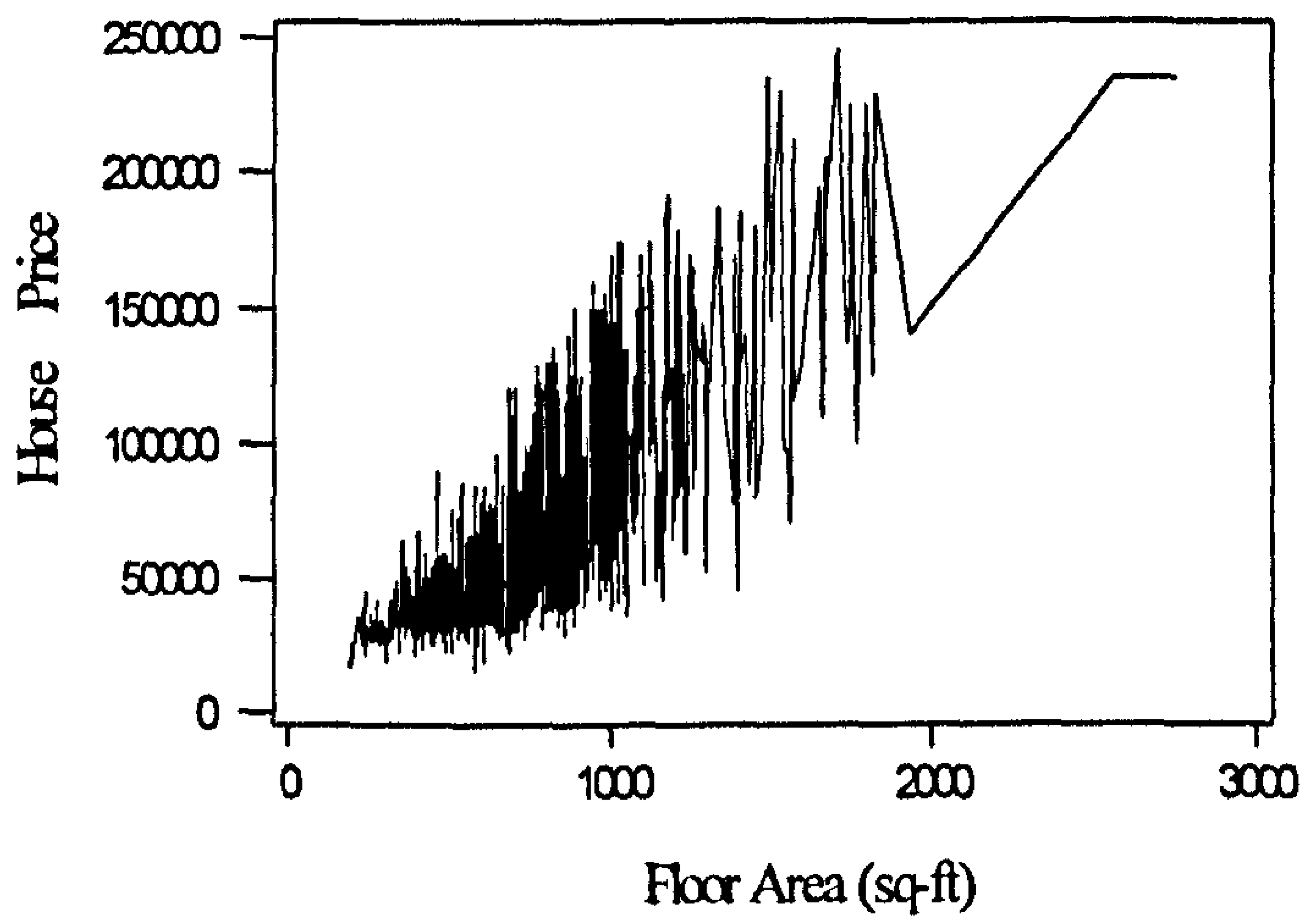


Figure 6.44

LOWESS Plot of House Price against Floor Area

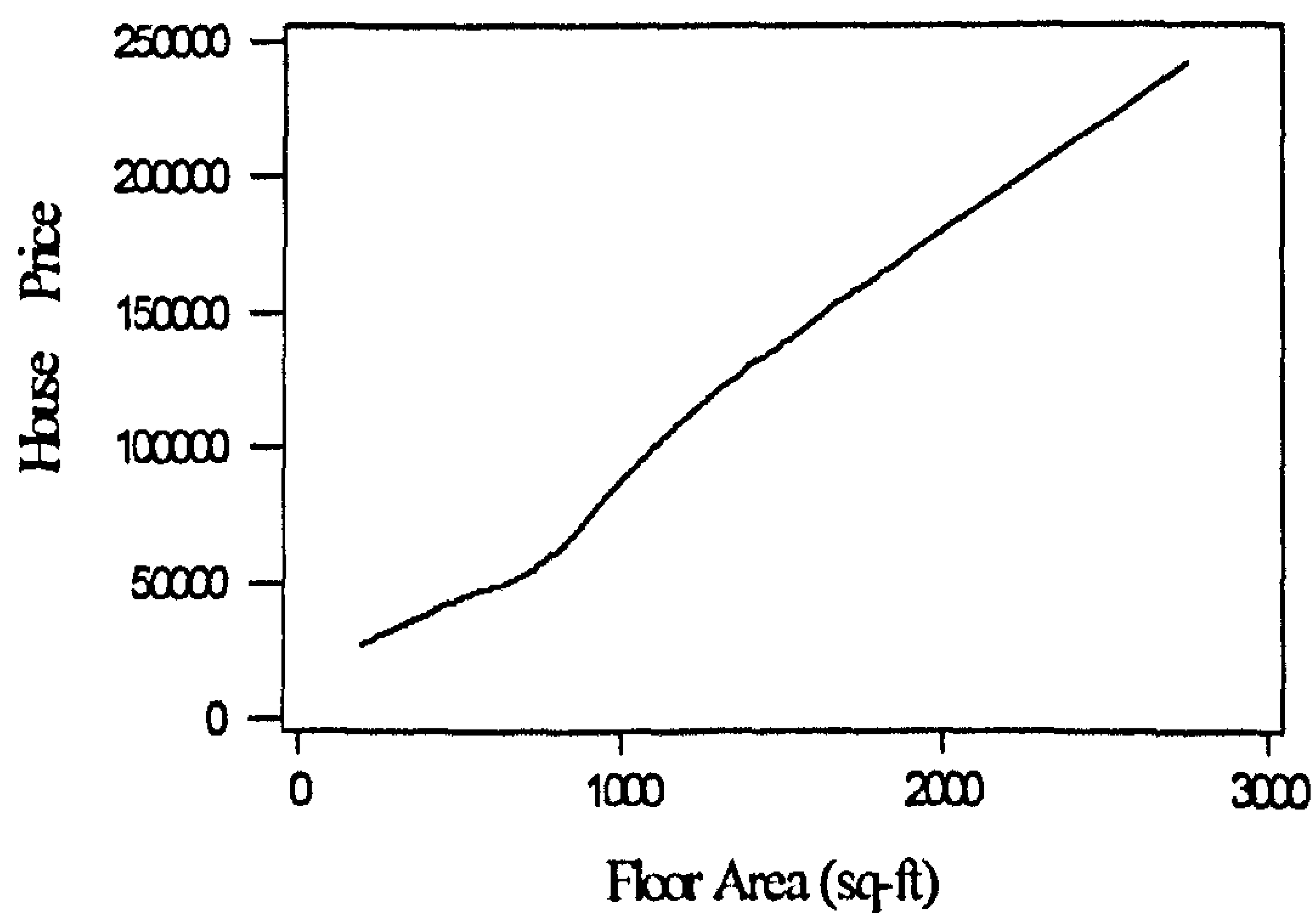
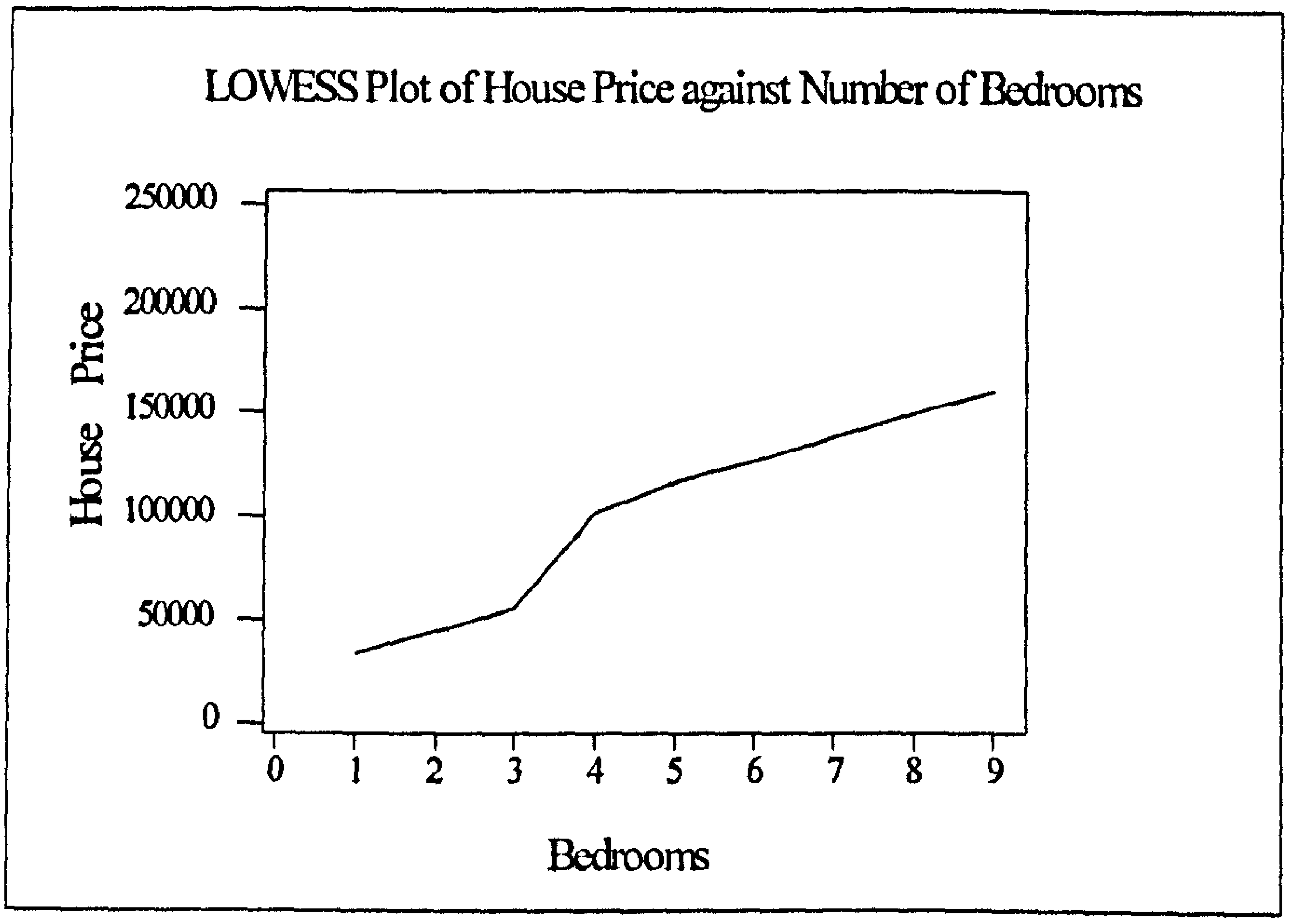
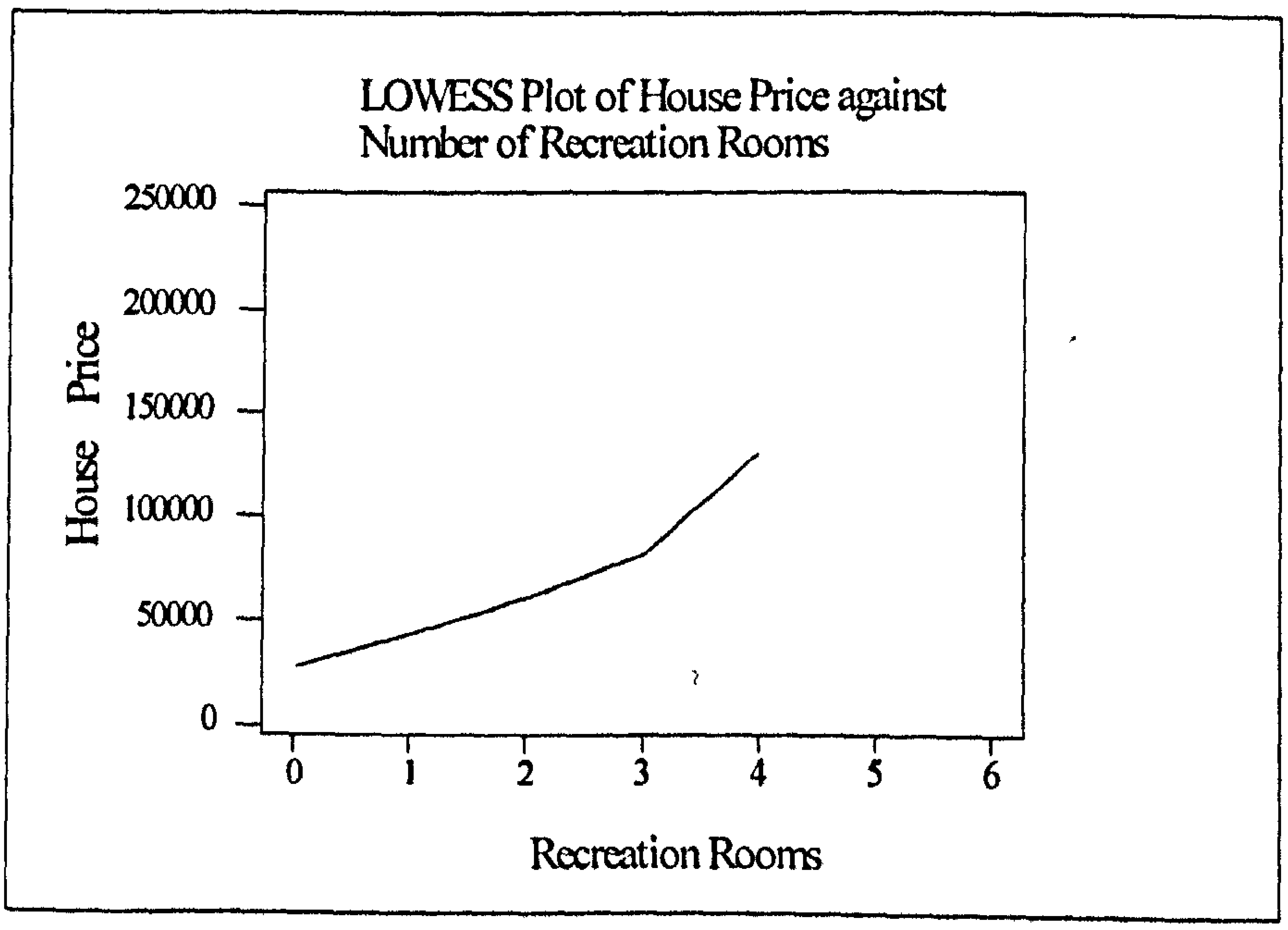


Figure 6.45


LOWESS Plot of House Price against Number of Bedrooms
House Price
250000
200000
150000
100000
50000
0
0
1
2
3
4
5
6
7
8
9
Bedrooms

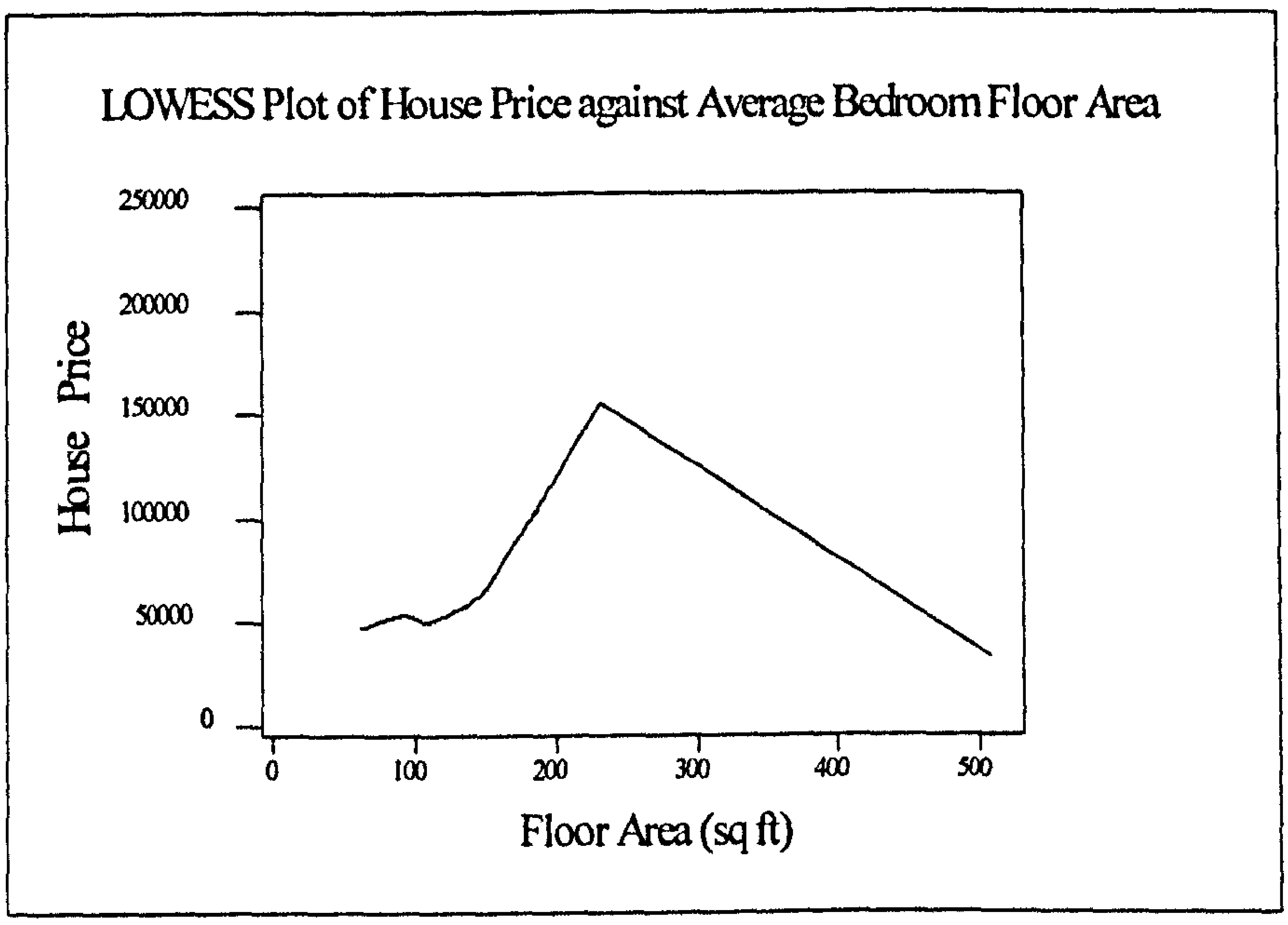

LOWESS Plot of House Price against
Number of Recreation Rooms
House Price
250000
200000
150000
100000
50000
0
0
1
2
3
4
5
6
Recreation Rooms

II. Bi-variate Plots against House Price

Simple plots of the dependent against the independent variables are useful in revealing the relationship between house price and housing attributes. In particular, they can help ascertain the correct functional form, point to evidence of possible heteroscedasticity and the presence of any unusual observations. However, care needs to be taken in their interpretation, since the multivariate nature of the hedonic model means they can be highly misleading (Chambers et al., 1983, pp. 258-259). It can therefore be useful to theorise the nature of the relation *a priori*, to prevent over-dependence upon the simple plot. Firstly, it can be anticipated that the relationship between house price and floor area will be non-linear, since the marginal demand for living space will not be constant, and thus the marginal increase in the cost of a unit of living space will increase at a decreasing rate. The same argument can be applied to the marginal increase in the number of rooms. For instance, the marginal willingness to pay for a second bathroom will be theoretically less than for the first bathroom. The relationship between house price and the locational attributes is less predictable though. It has previously been discussed that distance to the city centre will be negatively related and hypothetically logarithmic in nature, but the functional form of social class and housing quality is less of a certainty.

Figure 6.43 is the plot of floor area against house price. Despite the previous assertions, it can be seen that there is a general positive, linear relationship between the two variables, with house price increasing with floor area. However, the plot indicates two possible structural features. Firstly, the variation in house price between similar sized properties increases with floor area suggesting a heteroscedastic relation, with house price variation being less in smaller houses compared to larger houses. However, this could be due to omitted housing attributes, such as garden size and locational externalities, which may be more variable in larger houses. Secondly, there would appear to be a structural break at a floor area of around 850 sq-ft. At this point the relationship changes, and the variation in price and floor area becomes steeper. This break in slope is confirmed in Figure 6.44. This is a plot using a locally weighted scatterplot smoother (LOWESS), and is estimated by calculating new smoothed y-values (house prices) for each x-value (floor area) and then plotting a line between them. A LOWESS plot is useful in exploring the relationship between two variables without first trying to fit a specific model, such as a straight line. This plot confirms that the relationship between floor area and house price appears to be linear, and that a break in slope occurs at around 850 sq-ft. This could be caused by

Figure 6.46


LOWESS Plot of House Price against Average Bedroom Floor Area
House Price
250000
200000
150000
100000
50000
0
0
100
200
300
400
500
Floor Area (sq ft)

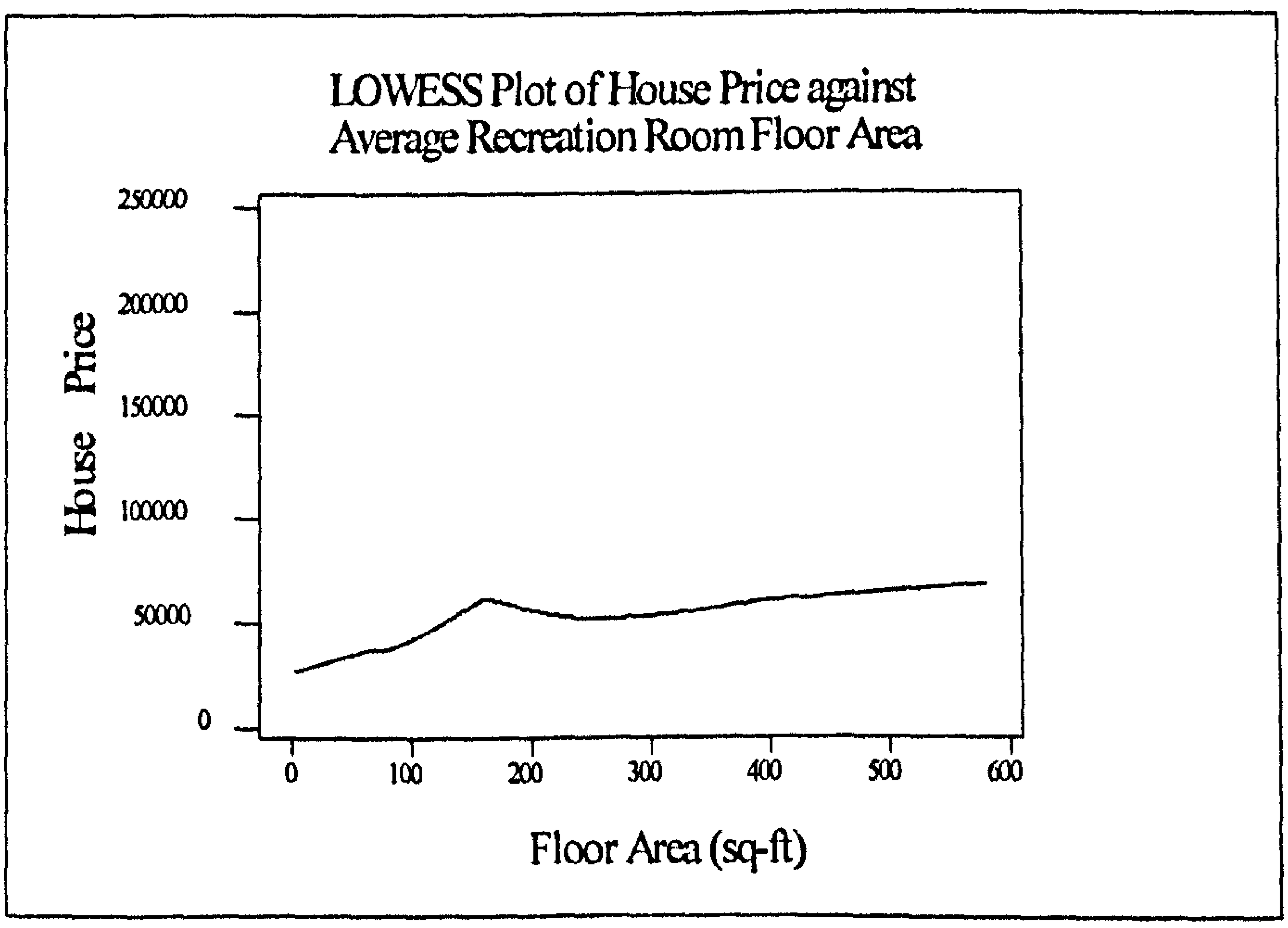
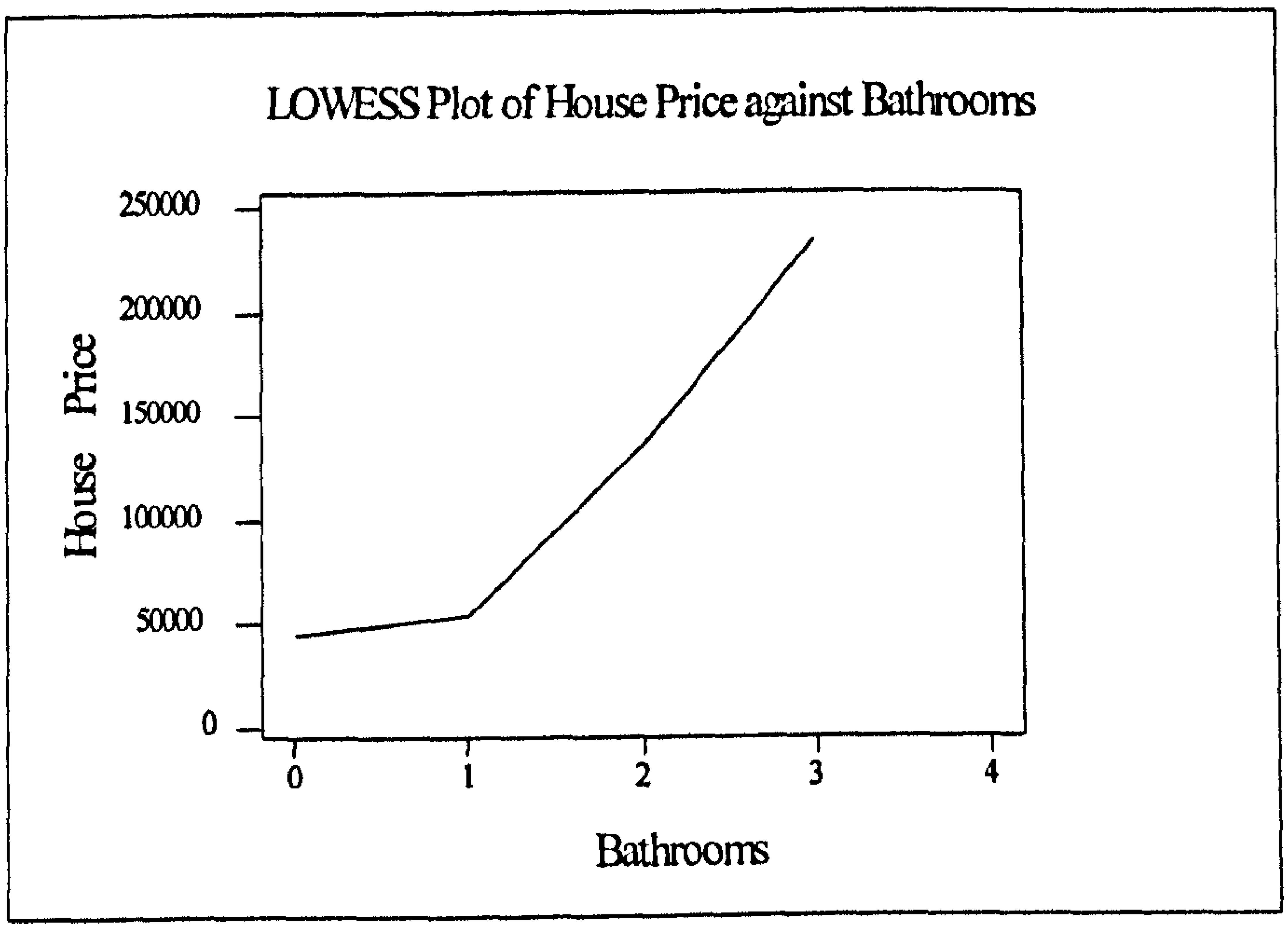
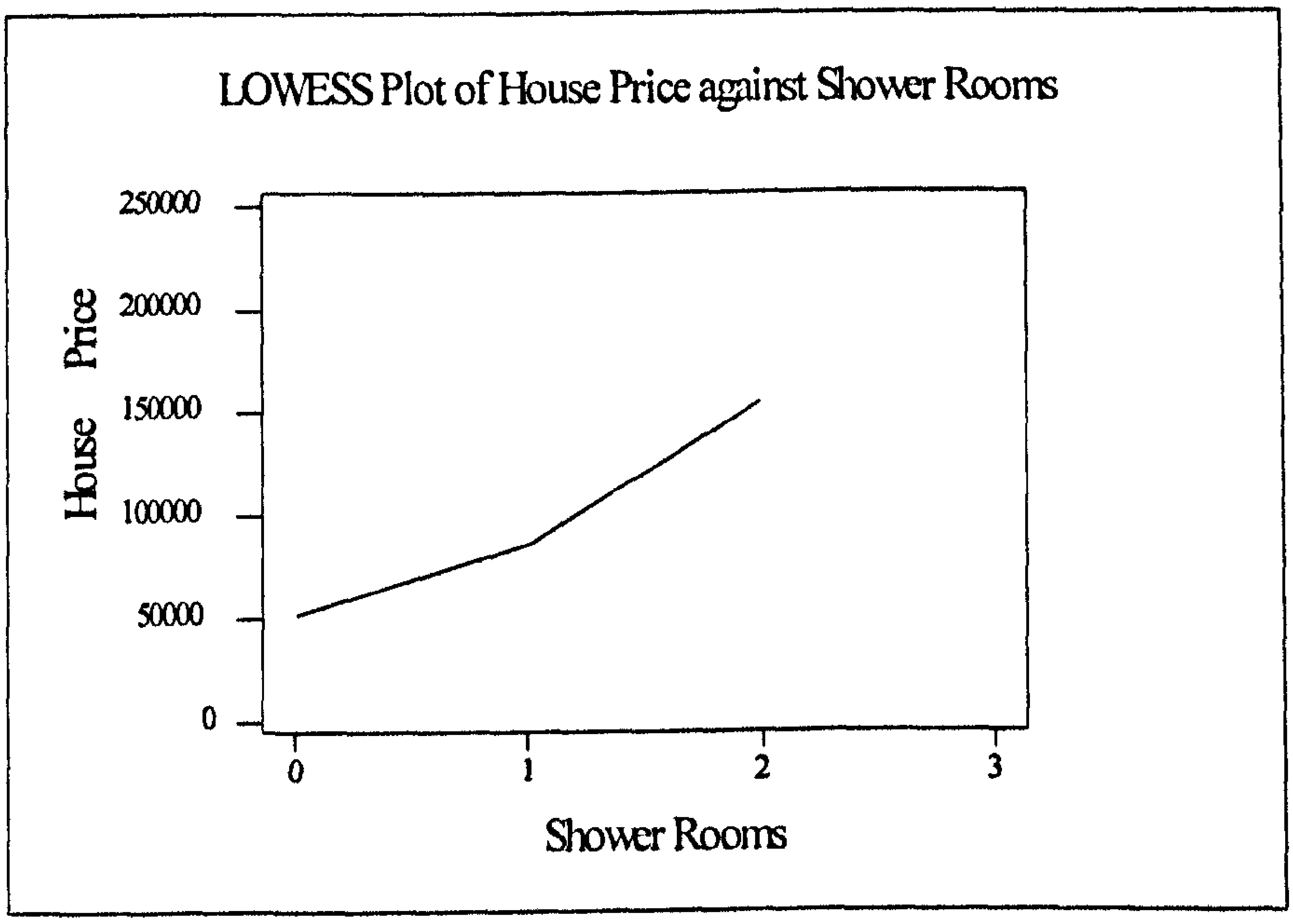

LOWESS Plot of House Price against
Average Recreation Room Floor Area
House Price
250000
200000
150000
100000
50000
0
0
100
200
300
400
500
600
Floor Area (sq-ft)

Figure 6.47


LOWESS Plot of House Price against Bathrooms
House Price
250000
200000
150000
100000
50000
0
0
1
2
3
4
Bathrooms


LOWESS Plot of House Price against Shower Rooms
House Price
250000
200000
150000
100000
50000
0
0
1
2
3
Shower Rooms

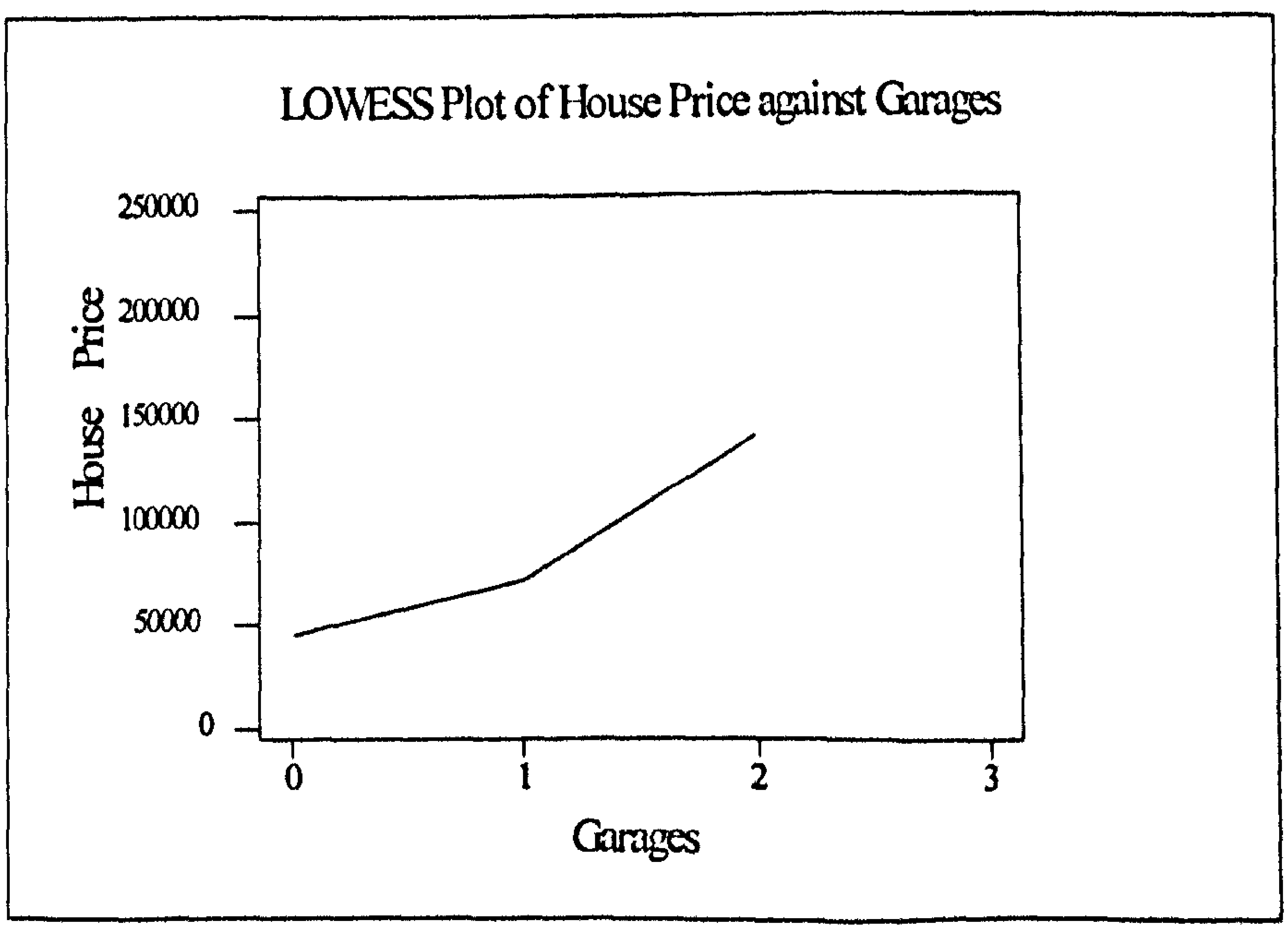
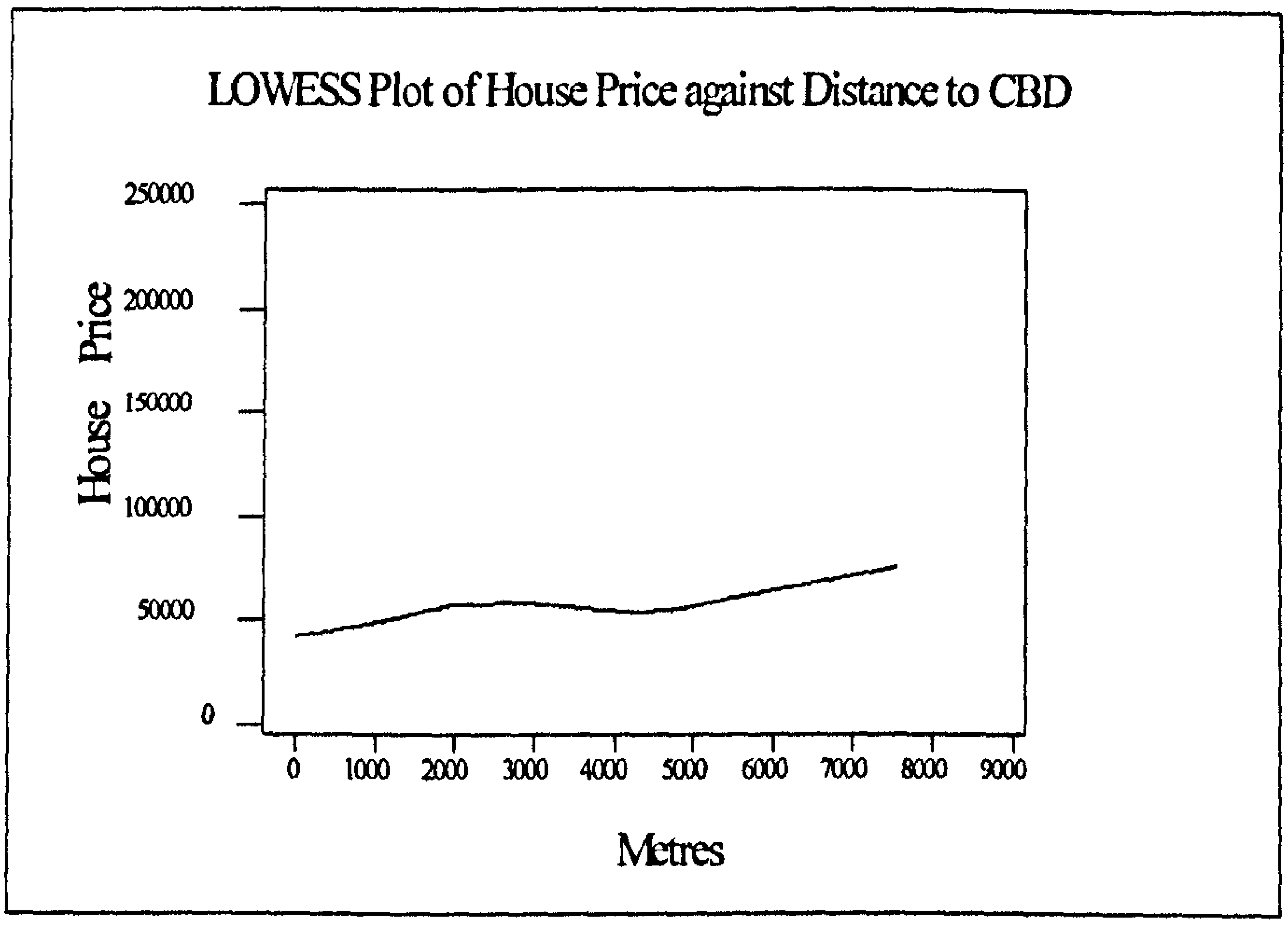
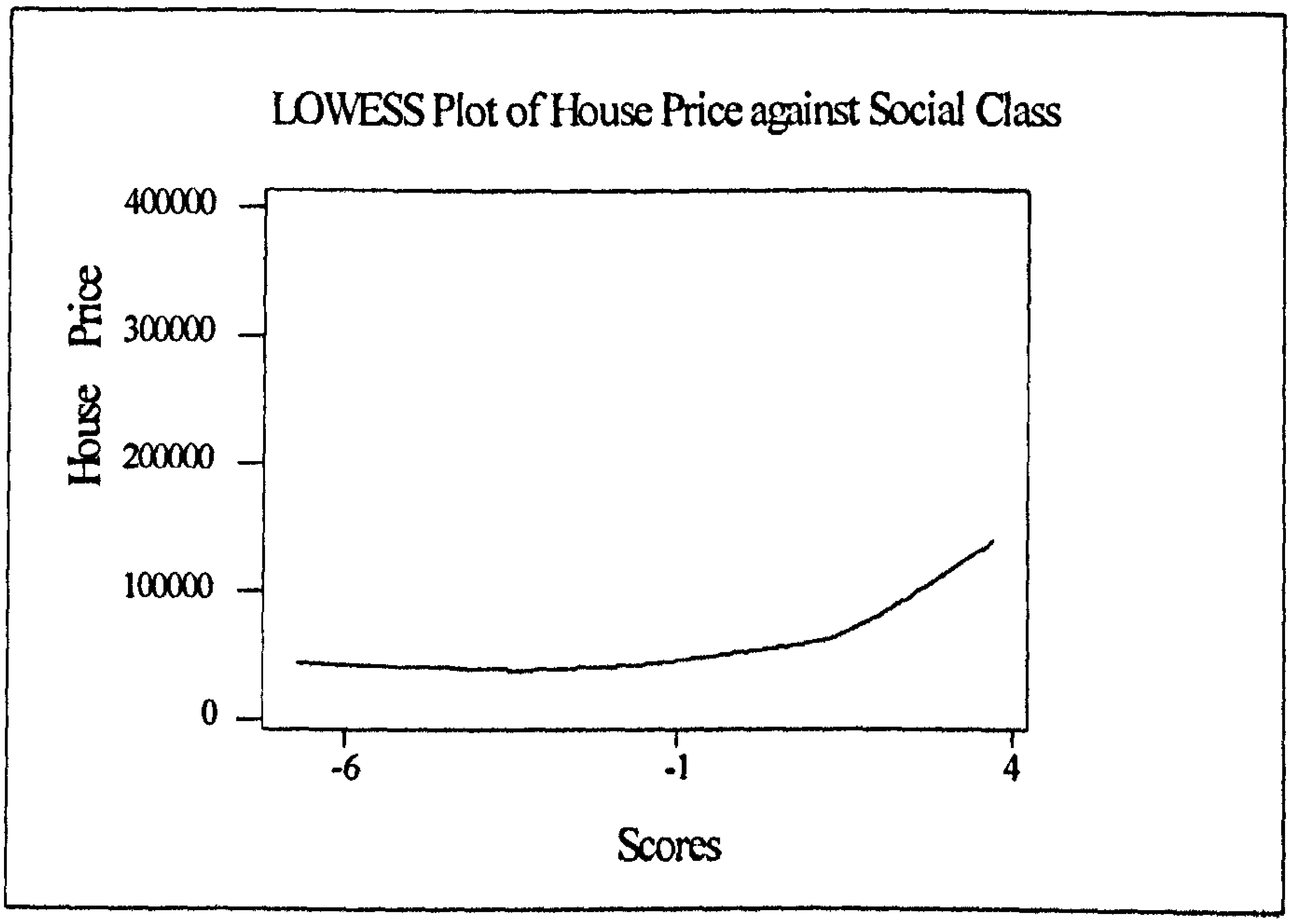
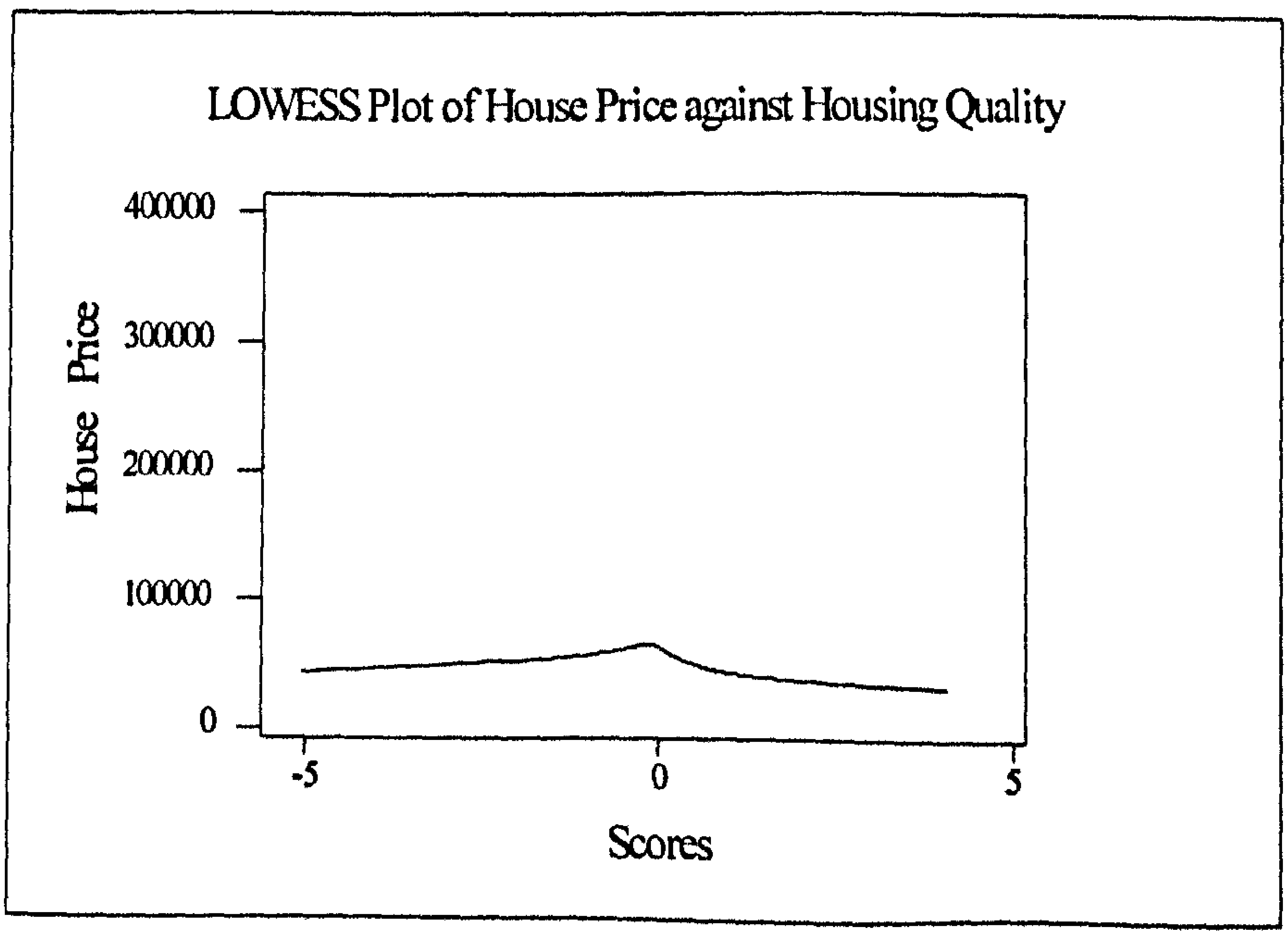

LOWESS Plot of House Price against Garages
House Price
250000
200000
150000
100000
50000
0
0
1
2
3
Garages

Figure 6.48


LOWESS Plot of House Price against Distance to CBD
House Price
250000
200000
150000
100000
50000
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Metres


LOWESS Plot of House Price against Social Class
House Price
400000
300000
200000
100000
0
-6
-1
4
Scores


LOWESS Plot of House Price against Housing Quality
House Price
400000
300000
200000
100000
0
-5
0
5
Scores

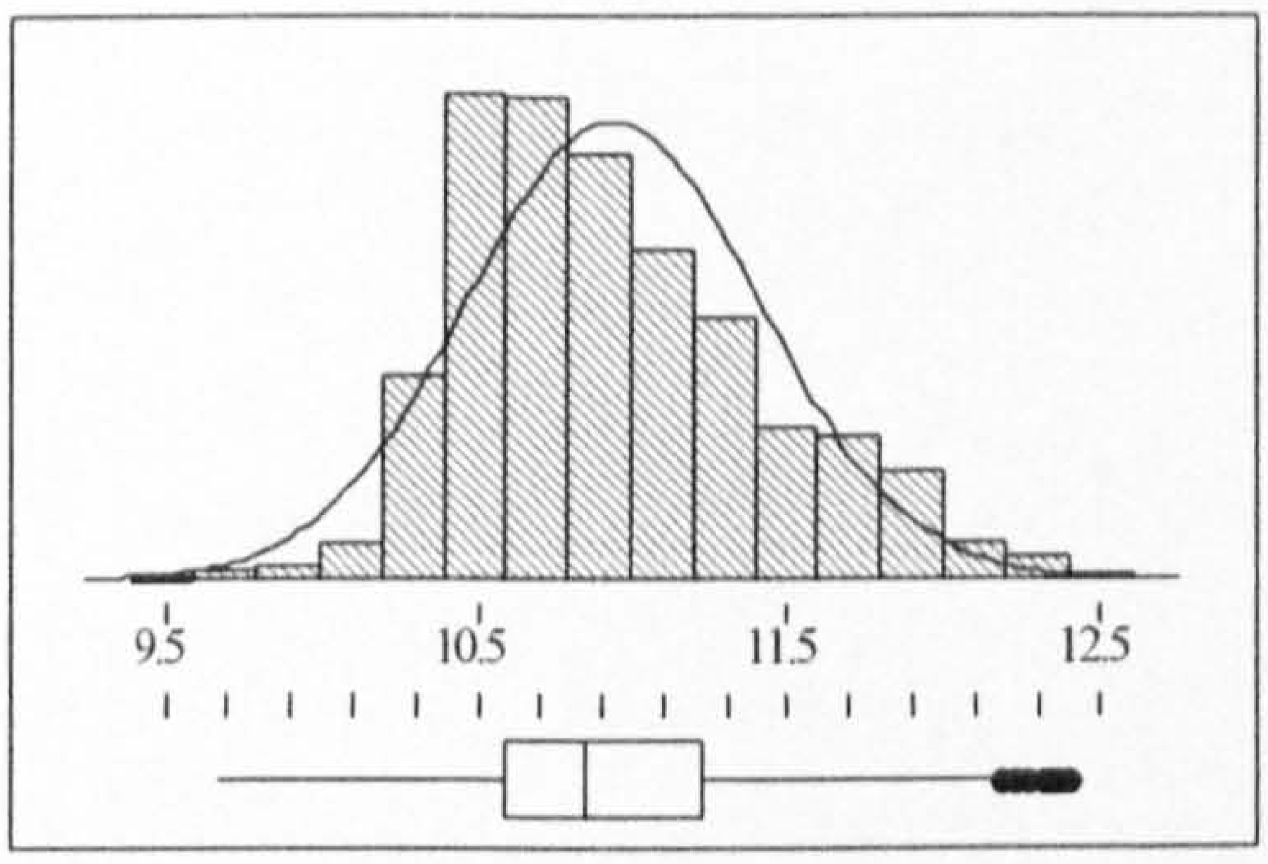
uncontrolled for housing attributes that are more prevalent in larger houses. It could also indicate two distinct functional relationships between house price and properties smaller and larger than 850 sq-ft. Similar plots for the number of bedrooms and recreation rooms against house price (Figure 6.45) also suggests the possible existence of a different functional relationship for larger properties, with distinctive breaks in the slope for properties with four or more bedrooms and recreation rooms. The functional forms again appear to be linear.

Figure 6.46 suggests that a non-linear relationship may exist between house price and average bedroom size. However, such a plot is also indicative of the influence of outliers illustrated in the box plot, and hence any assumption of non-linearity will have to be treated with care until further diagnostic tests have been performed. Figures 6.47 suggests that, when the attribute is present, a positive, linear relationship exists between house price and the number of bathrooms, shower rooms and garages. A linear relationship between distance to the city centre and house price is also implied in Figure 6.48, although the positive trend is counter-intuitive, an is probably caused by uncontrolled for attributes such as floor area increasing with distance. Finally, a non-linear relationship is suggested between house price and social class, although as was suggested in Figure 6.42, the change in slope coincides with several outlying observations. The influence of such observations are also evident in the plot for housing quality, with almost no relationship evident, except around the mean.

The simple bi-variate plots suggest that most of the independent variables have a linear relationship with house price, despite the assertions to the contrary, although there is perhaps evidence of more complex functional forms, particularly with respect to the floor area variables. There is also evidence that the outliers may have a disproportionate influence upon some of these relationships. However, what is evident is that the variance between house price and house size is not constant, but increase as house size increases. This may lead to subsequent modelling difficulties, specifically heteroscedasticity in the error terms leading to inefficient parameter estimates. These issues are investigated further in next section.

6.5.2.3 Transformation of the Dependent Variable

Transformation of the dependent variable is a common practice in many of the previous studies (e.g. Jones and Bullen, 1993) and is used to deal with the technical problems of non-linearity and variance heterogeneity. In particular, transformation of the data may be needed



Mean	10.94
Std Dev	0.47
Skewness	0.61
Kurtosis	-0.01
n of obs	1447
Min	9.68
1st Quartile	10.60
Median	10.86
3rd Quartile	11.22
Max	12.41

Figure 6.49

Natural Log of House Price

Figure 6.50

Plot of Transformed House Price against Floor Area

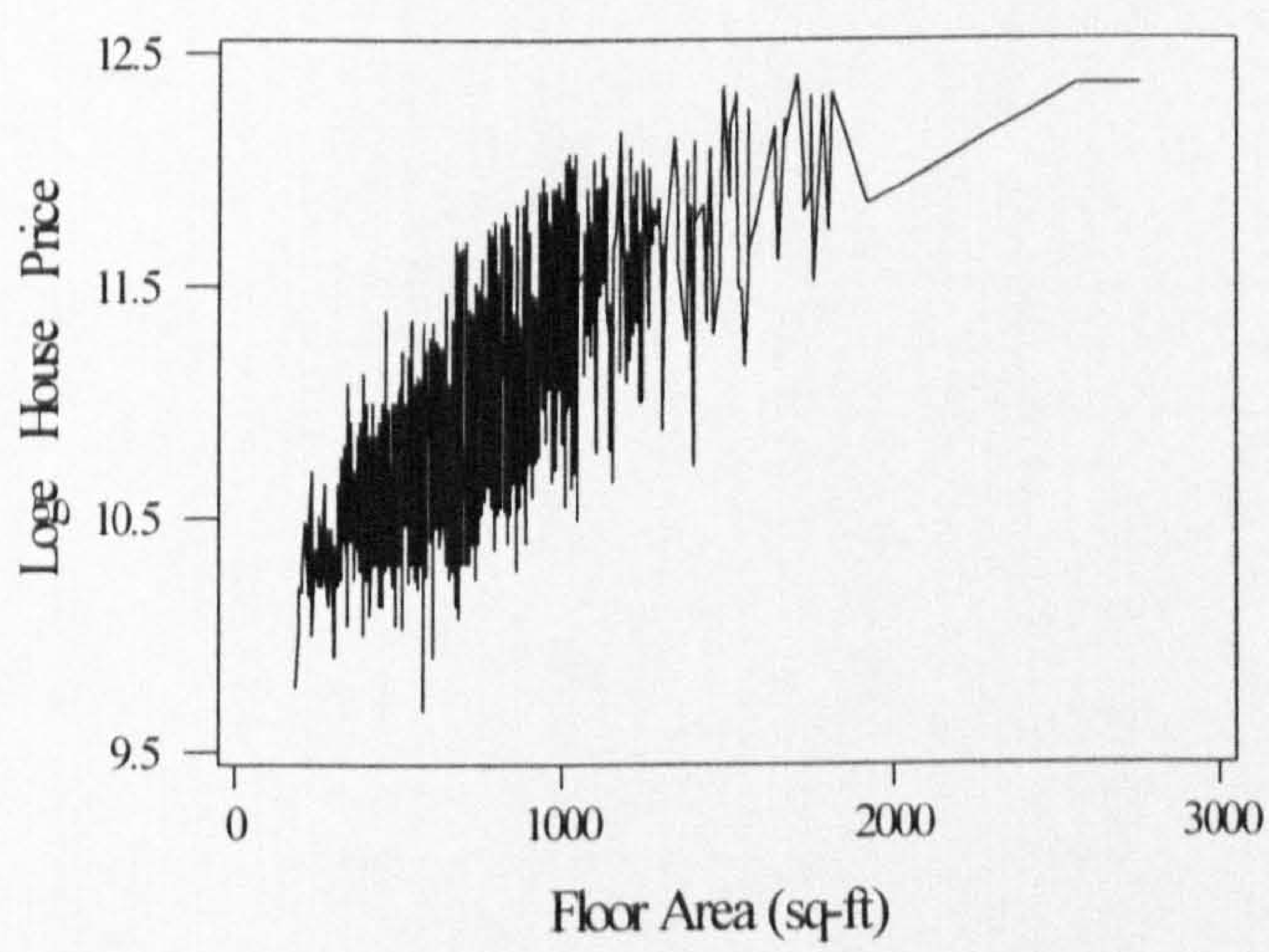
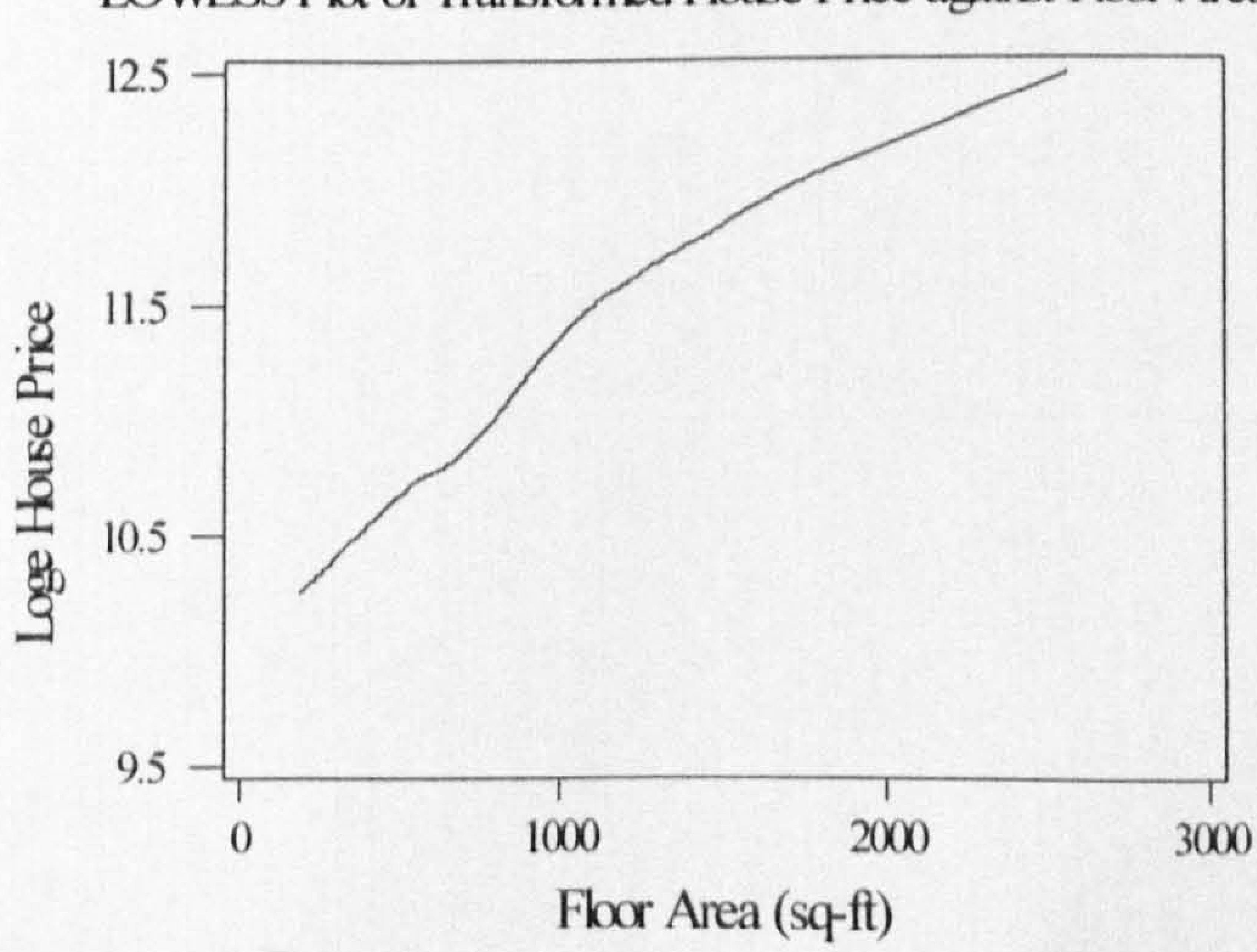


Figure 6.51

LOWESS Plot of Transformed House Price against Floor Area



to overcome non-linearity and variance problems related to house size (Addair et al. 1996). The above analysis has demonstrated that, although the relationship between house price and house size is linear, the variance of larger properties is greater than for smaller properties. Hence, to retain the functional relationship, but to reduce the effect of variance heterogeneity, the dependent variable was transformed using a natural log. The affect on the dependent variable can be seen graphically in Figure 6.49. The transformation has removed the skew and reduced the kurtosis, so the distribution is now virtually normal. The box plot indicates that several highly priced properties continue to remain as outliers though.

Figure 6.50 and 6.51 shows the relationship between the transformed house price variable and floor area. It can be seen that, although the functional relationship has in the main remained unchanged, the variance heterogeneity evident in Figure 6.43 has now been greatly reduced. Equivalent graphical analysis of the remaining structural attributes indicates similar, with the transformation of the dependent variable not significantly affecting the structural relationship, but reducing the variance heterogeneity.

6.5.2.4 Correlations Between the Housing Variables

Table 6.9 is a summary of the correlation coefficients between the variables. These coefficients can be used to identify strong correlations between the dependent and independent variables, and the presence of relationships between the independent variables. The first column shows the correlation coefficients between house price and the housing attribute variables. The values in bold indicate reasonably strong associations. The first thing to note is that the strongest association is with floor area, whilst the average room size variables only have mediocre relationships. The strong associations are indicative of larger properties, such as detached houses and the number of bathrooms and garages. This bias towards larger properties is also implied by the direction of the association of the house type dummy variables. The larger type properties, such as bungalows and semi-detached houses have a positive relationship with house price, whilst the small property types have negative associations. Counter-intuitive relationships occur with the age variables.

The remainder of the table summarises the pair-wise associations between a selection of independent variables. These are the only variables that have associations strong enough to present potential problems with multicollinearity. These collinearities were evident in Figures 6.36 and 6.37, and may become problematic in the hedonic model. It is striking that

Table 6.9.

Correlation Matrix of House Price and Selected Housing Attributes

Variables	*House Price*	Floor Area	Ave Bed Size	Ave Rec Size	Beds	Post-1964	1918-64	Pre-1918	Gdn: < 5m	Gdn: 5-50m	Gdn: > 50m
Floor Area	***0.762***										
Ave Bed Size	*0.157*	**0.507**									
Ave Rec Size	***0.458***	0.219	**0.571**								
Ave Kit Size	***0.419***	**0.475**	0.228	0.176							
Beds	***0.623***	**0.826**	0.153	0.087							
Recs	***0.558***	**0.727**	0.246	**0.625**	**0.623**						
ET	*-0.009*	0.167	0.140	-0.028	0.156						
MT	*-0.161*	0.084	0.084	-0.016	0.088						
SD	*0.058*	0.109	-0.067	0.025	0.150						
D	***0.566***	0.284	-0.017	0.003	0.333						
FPB	*-0.216*	-0.338	-0.036	-0.028	**-0.402**						
FCB	*-0.137*	-0.201	0.061	-0.047	-0.279						
M	*-0.101*	-0.109	0.001	0.047	-0.133						
B	*0.166*	0.030	0.119	0.060	-0.058						
EL	*-0.103*	-0.108	-0.126	-0.006	-0.056						
ML	*-0.147*	-0.127	-0.134	0.007	-0.072						
Baths	***0.418***	**0.407**	0.205	0.074	0.304						
Showers	*0.350*	0.312	0.121	0.020	0.289						
Full CH	*0.183*	0.016	-0.068	0.048	0.040						
Part CH	*-0.041*	0.009	0.023	-0.003	-0.006						
Gas	*0.247*	0.181	-0.033	0.080	0.239						
Garage	***0.572***	0.337	0.062	0.081	0.348						
ORP	*0.255*	-0.040	-0.124	0.025	-0.036						
Age: New	*-0.022*	-0.050	0.012	0.034	-0.066						
Post-1964	*-0.034*	-0.279	-0.270	-0.033	-0.183						
1918-64	*0.157*	0.133	0.053	0.070	0.106	**-0.475**					
Pre-1918	*-0.101*	0.182	0.241	-0.025	0.102	**-0.652**	-0.357				
Gdn: None	*-0.266*	-0.382	0.018	-0.030	**-0.470**	0.247	-0.180	-0.107			
Gdn: <5m	*-0.217*	0.042	0.081	-0.001	0.037	-0.286	-0.318	**0.577**			
Gdn: 5-50m	*0.128*	0.043	-0.183	-0.027	0.181	0.144	0.279	-0.394	**-0.611**		
Gdn: >50m	***0.463***	0.329	0.159	0.083	0.211	-0.086	0.264	-0.136	-0.223	-0.273	
Cons	*0.107*	0.127	0.053	-0.015	0.115	-0.069	-0.005	0.078	0.013	0.014	0.069
N. Mods	*-0.114*	0.086	0.045	-0.076	0.109	-0.195	0.018	0.192	0.159	-0.055	-0.058
Swm Pool	*0.222*	0.136	0.087	0.045	0.074	0.015	0.014	-0.027	-0.049	-0.060	0.218
Dist CBD	*0.232*	0.002	-0.179	0.010	0.069	**0.481**	0.093	**-0.591**	**-0.492**	**0.432**	0.214
Social	***0.477***	0.162	-0.003	0.024	0.121	0.280	0.052	-0.342	-0.315	0.242	0.219
H. Qual	*-0.136*	-0.189	-0.217	0.022	-0.131	0.154	0.191	-0.328	-0.226	0.207	0.038
LA > 50%	*-0.169*	-0.096	-0.047	0.010	-0.041	0.030	0.169	-0.177	-0.113	0.089	0.024

the remainder of the variables (not shown) only have very weak associations, although most of these are statistically significant at the 5% level (For 1430 observations, r is statistically significant if it exceeds 0.052). Finally, it is also worth noting the strong relationships between the different house size variables in column two.

6.5.3 Building an Initial Hedonic Model

6.5.3.1 Introduction

The mechanics behind this process are based upon checking that the assumptions underlying the hedonic model are not violated. To recapitulate, the five key assumption are:

I. that the relationship between the dependent and independent variables is linear
II. that there is no severe multicollinearity between the independent variables
III.that the fitting procedure has not be unduly influenced by unusual observations
IV. that the errors are homoscedastic
V. that the errors are not autocorrelated

If any of these assumptions is violated, then the desired properties of the OLS estimates no longer hold, and action is needed to produce a satisfactory model. At this preliminary stage, the first three assumptions are of importance, since these can cause the errors to exhibit heteroscedasticity and autocorrelation if violated. The third assumption is necessary, since unusual observations can unduly influence the results of an OLS regression analysis. It is necessary to be able to conclude that the model's estimates are not solely dependent upon outliers in the data, which can be regarded as a combination of leverage and discrepancy effects (Fox, 1991). The former are effects caused by influential X-observations (housing attributes), whilst the latter are caused by data points that have large residuals, and hence are related to the dependent variable (house price).

A top-down regression building approach was under-taken, which starts by including all the independent variables, and then discarding those that do not have a significant role in determining house price variation. The main guide is usually whether the associated t-statistic is statistically significant, although a statistic measuring the influence of multi-collinearity was also consulted. These tests are briefly outlined below.

6.5.3.2 Statistical Tests

Although several standard statistical tests could be applied to the parameters of the models, two tests were of particular importance. Firstly, a t-statistic was calculated by dividing the estimated regression coefficient by its standard error. This statistic was then used to test the hypothesis that no relationship existed between house price and each housing attribute; that is, the implicit price of an attribute is not significantly different from zero. By comparing the calculated t-statistic to the t-distribution, the null hypothesis that the housing attribute is uninfluential in determining house prices can be rejected if it exceeds the critical value in magnitude. For large samples, such as the one used in this research, the critical value is 1.96 at a five percent level of significance.

The second test statistic was the variance inflation factor (VIF). This test statistic is used to detect the presence of multicollinearity in the model. As was discussed in *Chapter Three*, multicollinearity is problematic within ordinary least squares regression since it may affect the estimation of regression parameters. The inter-related nature of housing attribute data, and the correlation coefficients in Table 6.8 suggests that multicollinearity may be a problem for some variables. A high VIF suggests collinearity, and as a general rule of thumb for standardised data, a VIF > 10 indicates harmful collinearity (Chatterjee and Price, 1977). The general approach regarding multicollinearity has been not to be too concerned if the R^2 from the regression exceeds the R^2 of any of the independent variables regressed on the remaining independent variables, or if the t-statistics are all significant (Kennedy, 1985). If multicollinearity is problematic, several remedies have been suggested. One is to obtain more data, since a larger sample size would provide additional information, helping to reduce variances. Another popular means of avoiding multicollinearity is to omit one of the collinear variables. However, this can cause a specification error if a relevant variable is omitted, causing the parameter estimates of the remaining variables to be biased. A third case is to use principal components in the regression, since these are orthogonal and thus uncorrelated.

6.5.3.3 The Initial Hedonic Models

As was stated above, two distinct hedonic models were estimated, based upon the role of house size in the model. The first hypothesizes that total floor area and the number of bedrooms and recreation rooms are both significant, and should enter the model

Table 6.10

Model 6.1 Total Floor Area

Predictor	Coeff	St.Error	T-stat	VIF
Constant	61350	113	539.19	
Floor Area	34.29	2.390	14.35	5.2
ET	-1334	1552	-0.86	4.3
SD	1932	863	2.24	6.0
D	20310	2261	8.98	4.9
FPB	1790	2452	0.73	6.7
FCB	220	2200	0.10	2.7
M	-1416	2529	-0.56	2.2
B	19444	2627	7.40	2.3
EL	-962	2048	-0.47	2.4
ML	-3090	1796	-1.72	3.0
Bedrooms	-1587	969	-1.64	4.5
Rec Rooms	600	902	0.67	2.3
Baths	7046	1492	4.72	1.4
Showers	5752	1198	4.80	1.4
Full CH	5364	1563	3.43	2.8
Part CH	245	2458	0.10	1.4
Gas	-864	1700	-0.51	2.5
Garage	4115	788	5.22	1.8
ORP	4169	906	4.60	1.8
Age: New	2964	3705	0.80	1.2
Post 1964	-219	1465	-0.15	5.0
1918-64	1937	1468	1.32	3.4
Gdn: None	-3200	646	-4.95	7.3
Gdn: <5m	3579	1177	3.04	6.3
Gdn: 5-50m	7554	1833	4.12	4.1
Con	2333	2040	1.14	1.0
Needs Mods	-7073	1571	-4.50	1.2
Swm Pool	3554	4936	0.72	1.1
Dist CBD	-2.27	0.277	-8.18	3.0
Social	5600	307	18.22	2.7
H.Qual	-1675	452	-3.70	1.7
LA > 50%	3902	1913	2.04	2.0
s	17160		R-sq(adj)	82.5

independently of one another. The second hypothesizes that it is not the total floor area, but the average floor area of the habitable rooms that are significant. However, this model also assumes that the number of habitable rooms will be important, given the supply and demand mechanism outlined in *Chapter One*. Hence this model incorporates the average bedroom, recreation room and kitchen floor areas, and the number of bedrooms and recreation rooms. The functional forms used in each model were based upon those implied by the previous bi-variate plots. The suitability of these functional forms will be tested during the model building process, and adjusted accordingly. To prevent rounding errors due to large numbers, and to facilitate interpretation of the models, the continuous independent variables were deviate around their means, so the models were estimated with regards to the stereotypical property of a three bedroomed, two reception roomed mid-terrace house with one bathroom, no shower room and no garage and a garden between 0 -5 metres in length. In addition, since the natural logarithm of house price is being used, the subsequent estimates will be in terms of percentage increases, providing that these estimates are relatively small (below 0.25 in absolute terms for Tufte, 1974). However, relatively large positive and negative values are poor approximations, (Jones and Bullen, 1993), and it is therefore sensible to appreciate the size of the effects by transforming the logarithms and expressing them as the difference over the base price of the stereotypical property. This has the effect of presenting the estimated implicit price in pounds.

I. Model 6.1- The Total Floor Area Model.

Table 6.10 is a summary of model 6.1, which was estimated using the full set of structural variables and total floor area as a measure of house size. The number of bedrooms and recreation rooms were also included, since these may have an influence on house price, independent of property size. For instance, two identically sized properties may have different numbers of bedrooms and thus may sell for a different price. The four columns consist of the estimated regression coefficient, which represent the implicit prices of the attributes, the associated standard error, the calculated t-statistic and the variance inflation factors. The latter two are of interest since these can be used to determine whether the variable is included in the model. The first thing to note is the insignificance of many of the variables. In particular, a third of the dummy house type variables and the number of bedrooms and recreation rooms fail to be significant at the five percent level. Table 6.9 shows that although both these latter two variables had a moderately strong relationships with house price (0.623 and 0.558 respectively), they were very strongly related to floor

Table 6.11

Model 6.2 Average Floor Area

Predictor	Coeff	St.Error	T-stat	VIF
Constant	59720	121	489.68	
Ave Bed Floor	81.24	48.07	1.69	9.6
Ave Rec Floor	58.02	6.49	8.94	6.4
Ave Kit Floor	52.00	8.43	6.17	1.3
ET	-1889	1628	-1.16	4.3
SD	2355	920	2.56	6.9
D	19249	2350	8.19	5.7
FPB	2988	2644	1.13	6.7
FCB	-175	2513	-0.07	2.4
M	-718	2660	-0.27	2.6
B	16951	2801	6.05	2.3
EL	-584	2163	-0.27	2.5
ML	-1381	1917	-0.72	3.2
Bedrooms	8092	632	12.80	2.7
Rec.Rooms	9532	6150	1.55	1.4
Baths	8989	1523	5.90	1.4
Showers	5836	1195	4.88	1.4
Full CH	6995	1685	4.15	2.8
Part CH	1881	2442	0.77	1.5
Gas	-655	1259	-0.52	2.5
Garage	3432	815	4.21	1.9
ORP	4451	967	4.60	1.9
Age: New	8385	7487	1.12	1.2
Post 1964	-1910	1480	-1.29	5.0
1918-64	532	1520	0.35	3.4
Gdn: None	-1480	353	-4.19	6.7
Gdn: <5m	3464	1237	2.80	5.9
Gdn: 5-50m	7479	1860	4.02	3.8
Cons	2970	2750	1.08	1.0
Needs Mods	-5978	1742	-3.43	1.2
Swm Pool	6179	4827	1.28	1.1
Dist CBD	-1.83	0.28	-6.4	3.1
Social	5407	318	16.96	2.7
H.Qual	-1648	493	-3.34	1.7
LA > 50%	2415	1145	2.11	1.9
s	17980		R-sq(adj)	81.1

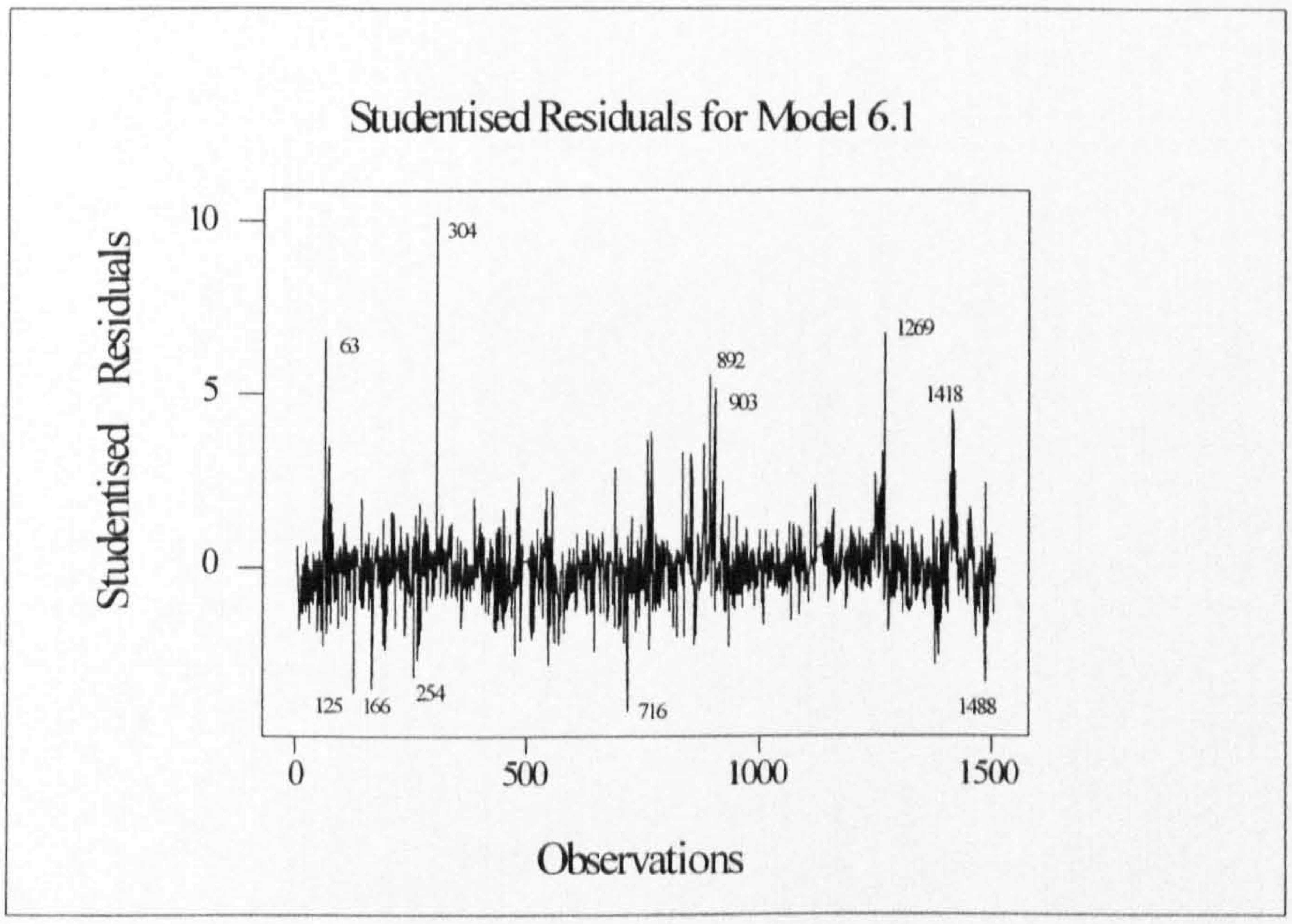
area (0.826 and 0.727 respectively). This collinearity may explain their insignificance in the model, even though their VIFs are not particularly significant. Further to this, the number of bedrooms also has a counter-intuitive sign indicating that price declines as the number of bedrooms increase, corroborating that multicollinearity in the model may be problematic. However, it may also reflect the fact that a lot of Victorian properties tend to have many small rooms that are often 'knocked through' in improved properties. However, such results may also be caused by unusual observations in the model having an anomalous effect. These will be examined further in the next section

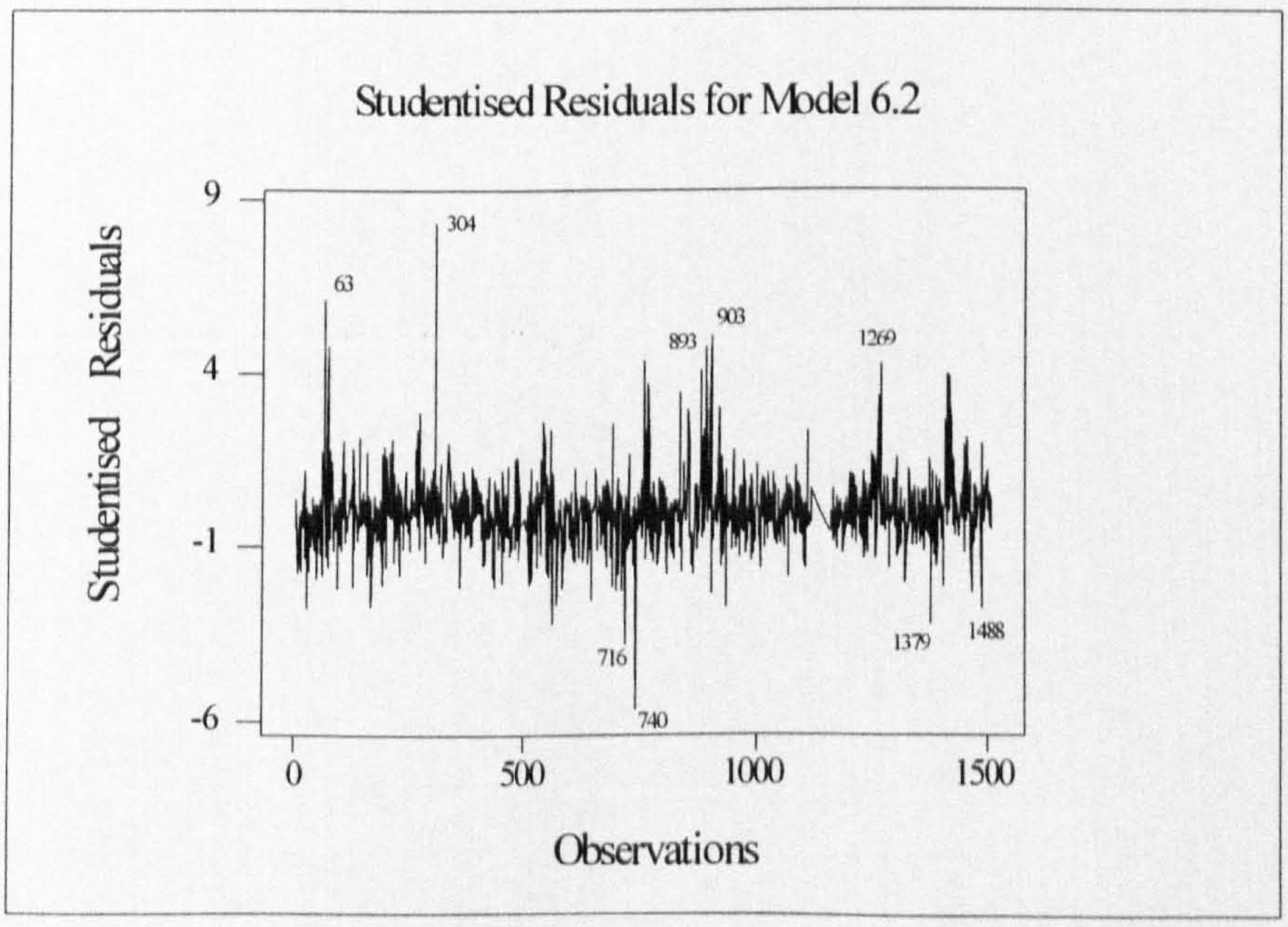
The other set of variables to take note of are the age variables. Their insignificance is unusual, since theoretically the age of the property should have a tangible influence on the price. However, their correlation coefficients with house price are very small, and somewhat counter-intuitive, whilst their associations with other independent variables, such as distance to CBD, are quite strong. This implies that correlation with locational attributes may be problematic, as was discussed in detail in *Chapter Three* and illustrated in the work by Heikkila et al, (1989). Such collinearity may not be picked up by conventional statistical tests, and as Figure 6.37 demonstrates, the age of properties do vary systematically with distance from the city centre. Due to the nature of the coarse age bands and lack of information, this problem of collinearity cannot be adequately dealt with by simply separating out age and distance, or adding improved age data. The remaining insignificant variables, such as conservatory, can be ascribed to the small number of observations associated with these attributes. The remaining variables were significant and did not display any adverse signs of multicollinearity. The R-squared (adjusted) statistic suggests that the model accounts for just over four fifths of house price variation.

II. Model 6.2. The Average Floor Area Model

In a manner similar to the previous model building exercise, a new model was estimated using the average room sizes as a measure of floor area - Table 6.11. The notable result is the insignificance of average bedroom room floor area, and the strength of the number of bedrooms, in contradiction to model 6.1. This implies that the number of bedrooms, and not necessarily their size, is important in house price determination. This seems not to be the case with respect to recreation rooms, and can be explained to some extent by the housing market supply and demand mechanisms outline in *Chapter One*. However, two other factors could also explain the lack of significance. Firstly, as Figure 6.5 illustrated, there would

Figure 6.52


Studentised Residuals for Model 6.1
Studentised Residuals
10
5
0
304
63
1269
892
903
1418
125
166
254
716
1488
0
500
1000
1500
Observations


Studentised Residuals for Model 6.2
Studentised Residuals
9
4
-1
-6
304
63
903
893
1269
716
740
1379
1488
0
500
1000
1500
Observations

appear to be very little variation in average bedroom floor area between properties, and secondly, the variance inflation factor suggests that multicollinearity with average bedroom size could be problematic. The remaining parameter estimates are similar to those in Model 6.1. The R-squared (adjusted) statistic is also lower in this model.

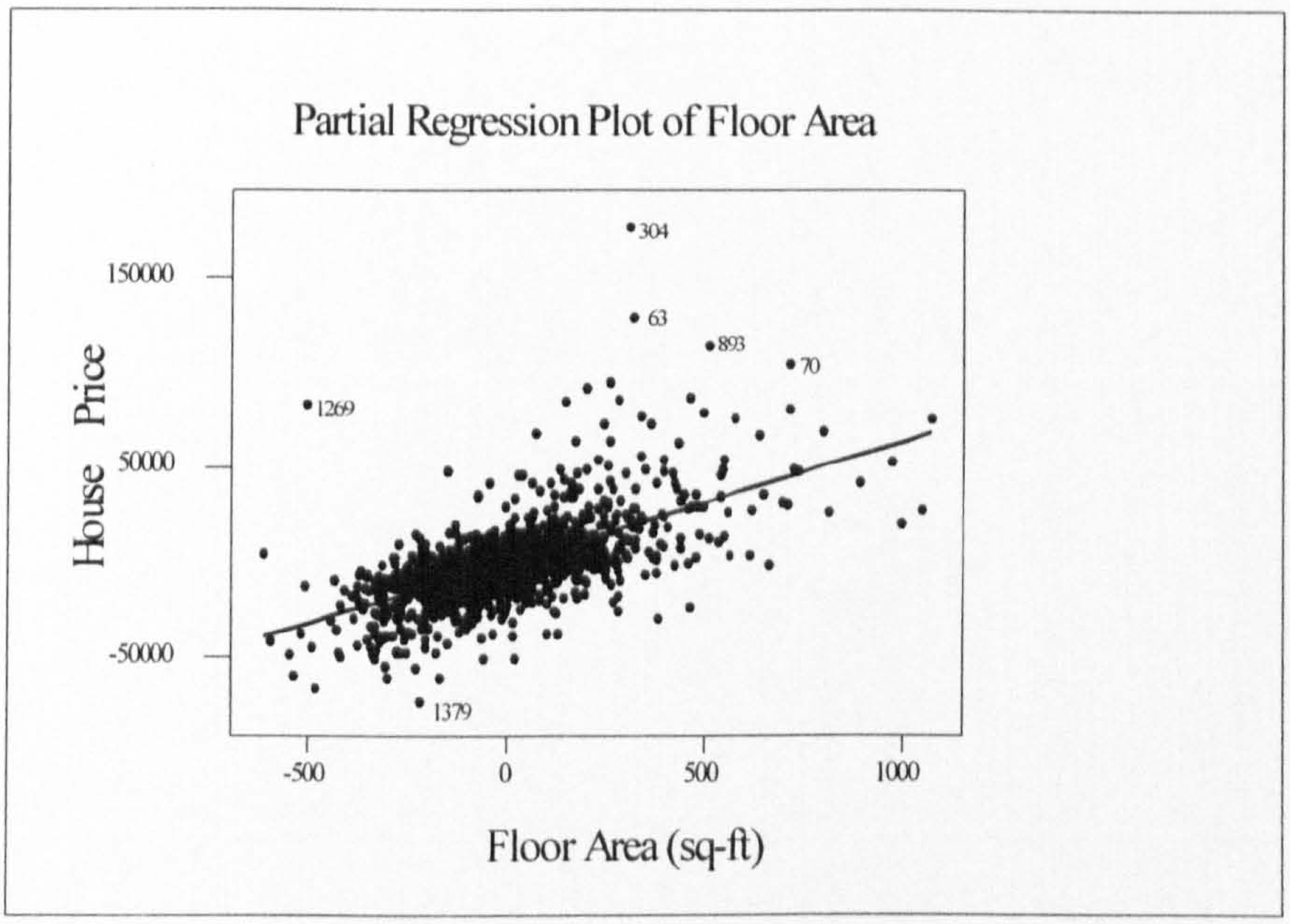
6.5.3.4 Diagnostic Tests

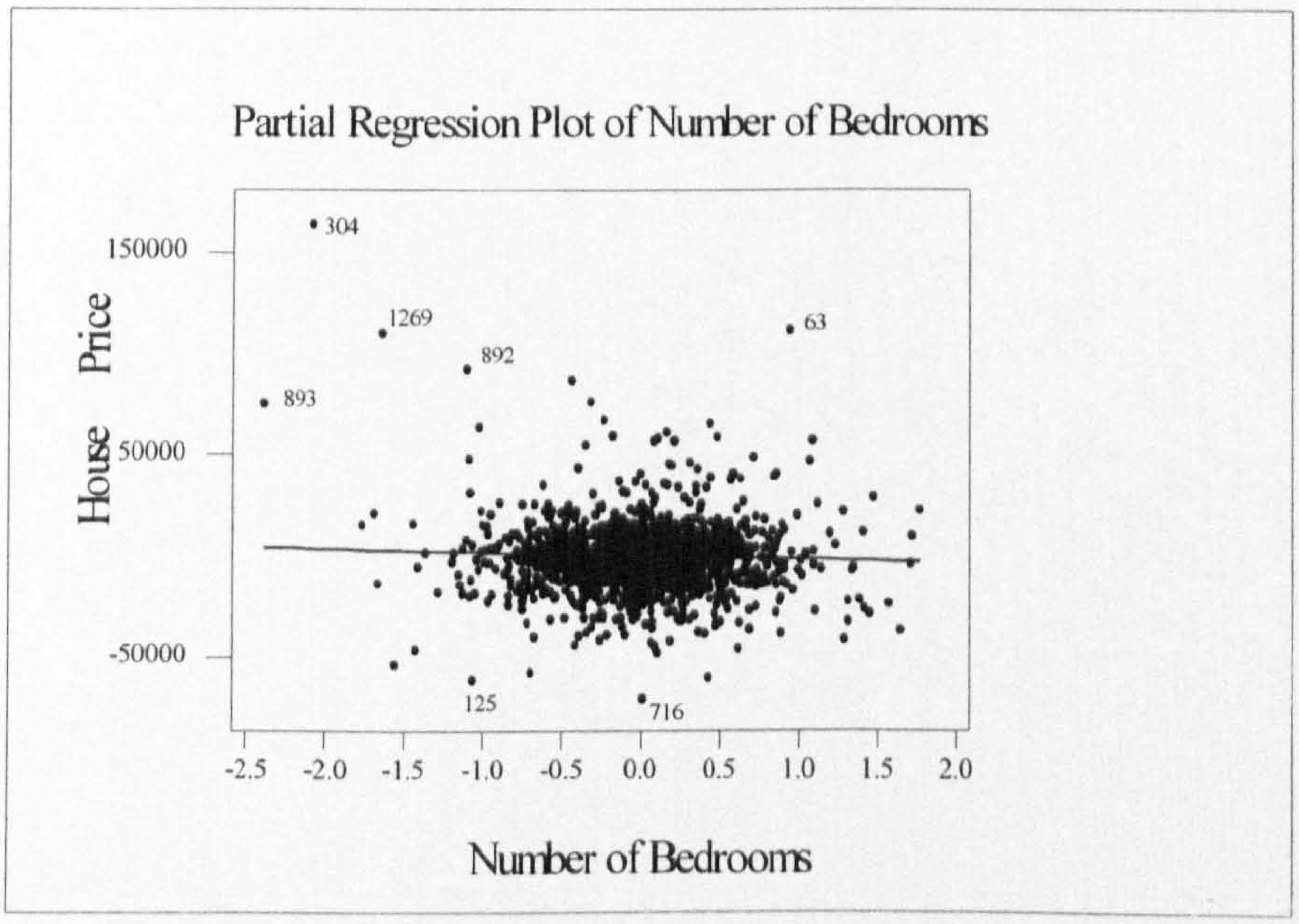
Before any of the insignificant independent variables were dropped from the model, diagnostic tests were performed on the model to avoid reliance on the statistical summaries. As discussed in *Chapter Two*, graphical diagnostics are an important, integral part of the model building process, and are among the most sophisticated diagnostical techniques available (Dunn, 1989). In particular, the models need to be checked for non-linearities, heteroscedasticity and unusual observations that may have an anomalous influence upon the regression parameters. Three principle diagnostic tests were performed. Firstly, the residuals from both models were examined to check for unusually large values; secondly, partial regression plots for each of the independent variables were analysed; and finally, joint regression diagnostic tests were performed upon the partial regression residuals.

I. Studentised Residuals

Relatively large residuals may indicate that a particular observation may be influential. But an influential data point can also be associated with a small residual. Because of this, it is useful to consider the residual that is obtained for each observation when the regression is estimated with that particular observation omitted. These are called studentised residuals, and assuming they follow a normal distribution, any residual greater than the critical value of 1.96 (a t-distribution at five percent significance level), can be regarded as an outlier and should receive special attention. Figure 6.52 are plots of the studentised residuals for Models 6.1 and 6.2 respectively. The first thing to note is that anonymously large residuals are estimated in both models, particularly positive residuals. Positive residuals suggest that the model under-estimates house price for a particular observation. A number of outliers have been identified, and it would appear that the same observations are influential in both models. This may suggest that either these properties have unusual attributes, or are in unusual locations, for the asking price.

Figure 6.53: Partial Regression Plots for Model 6.1





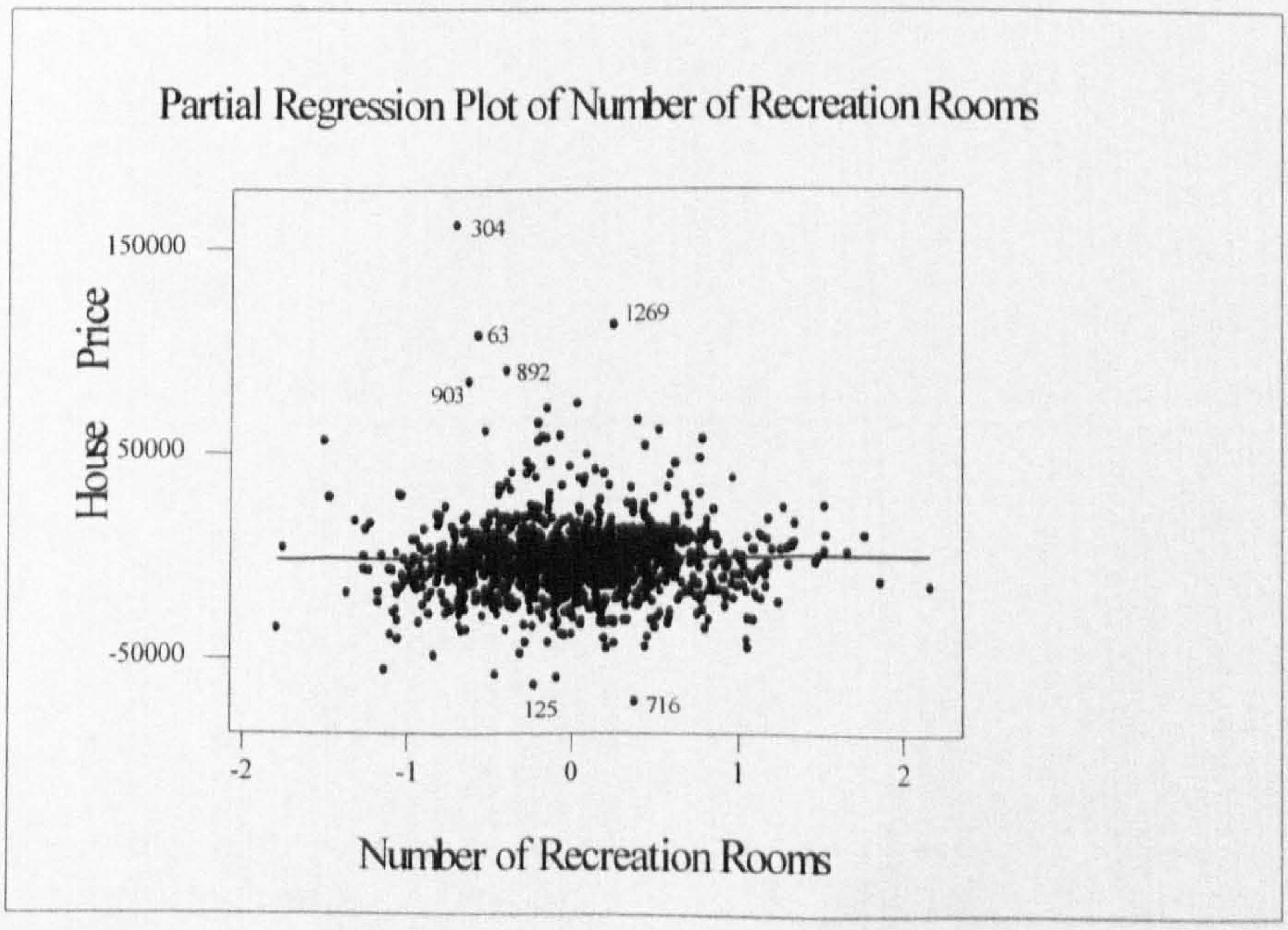
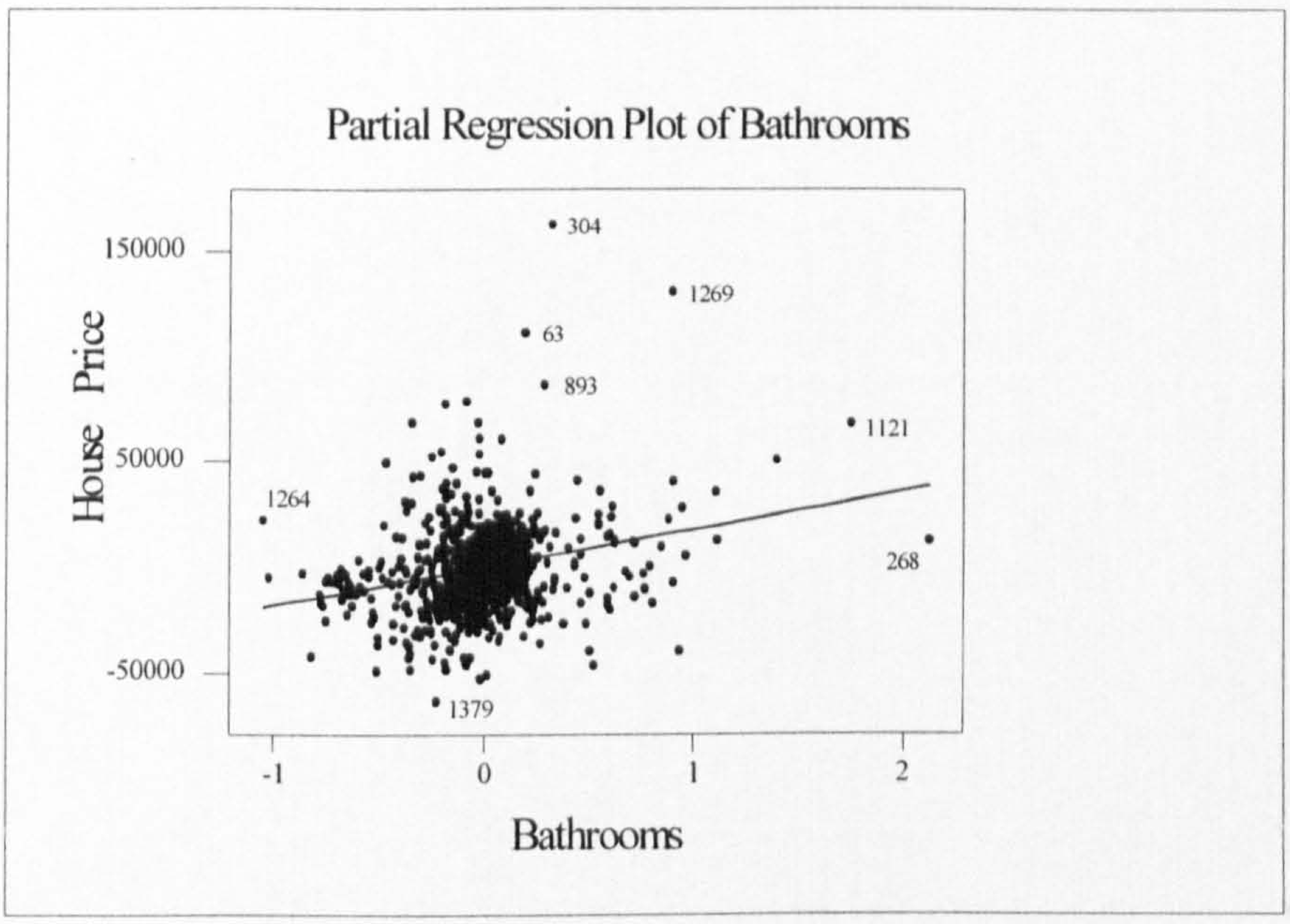
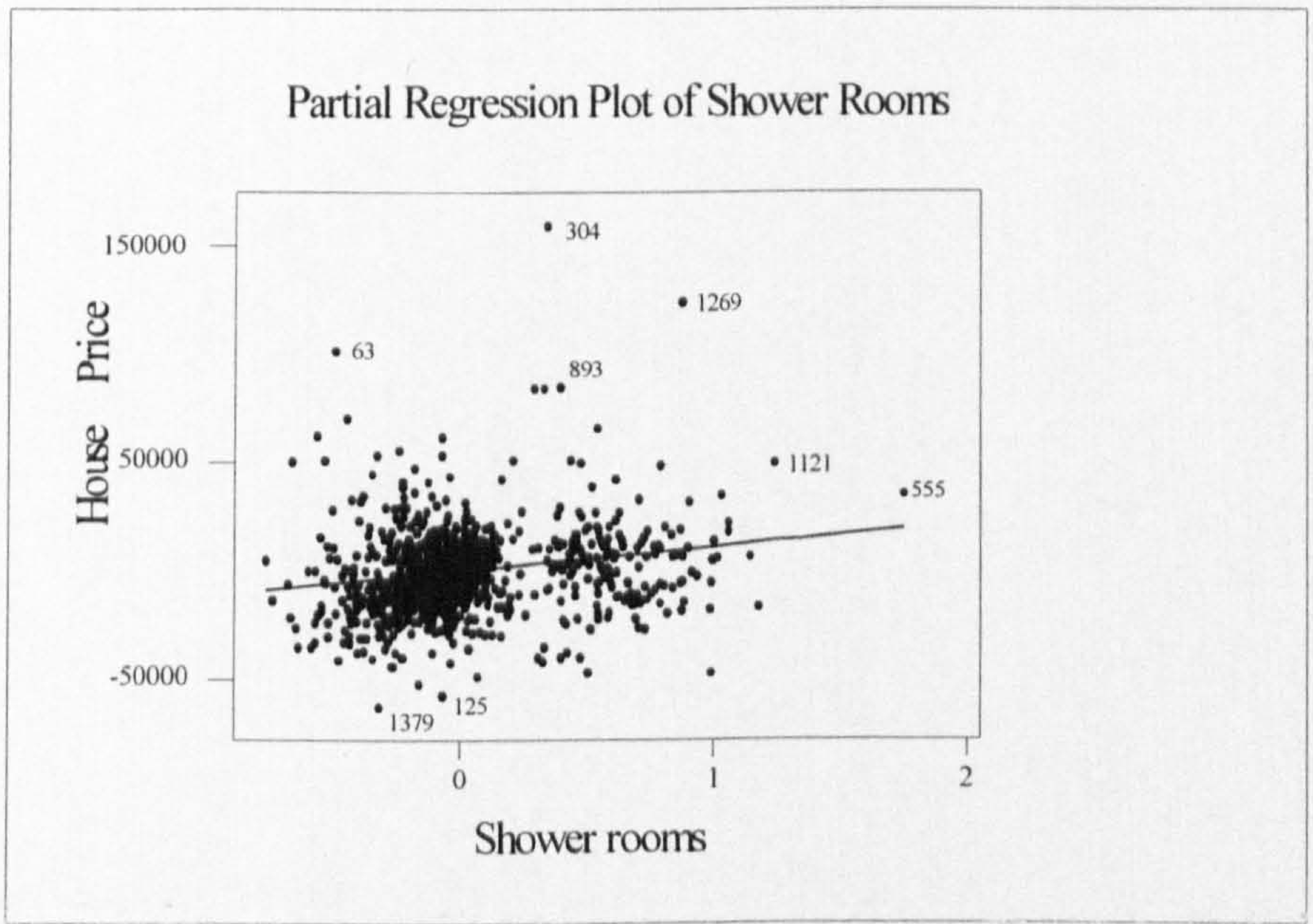


Figure 6.53 (cont.)





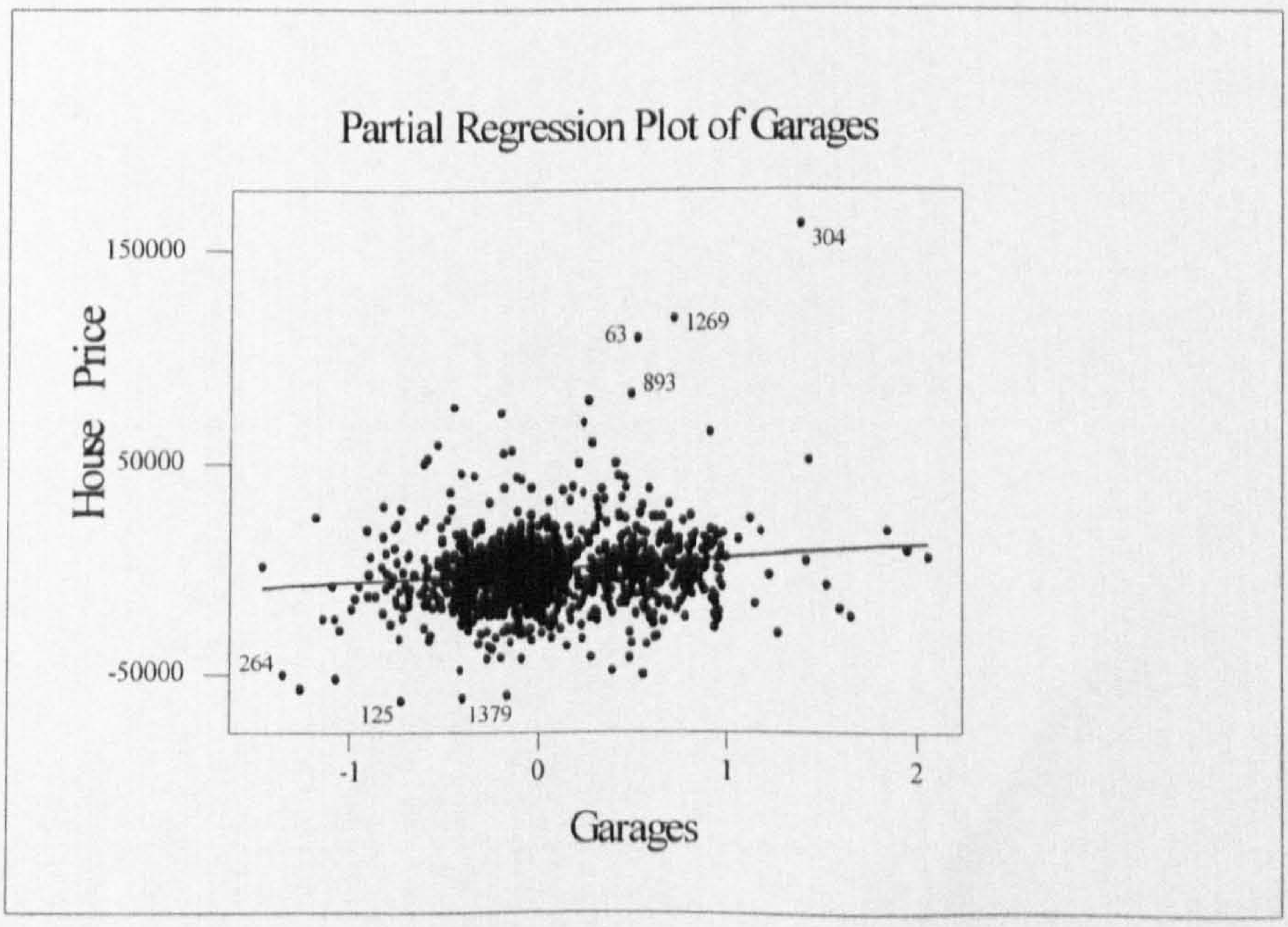
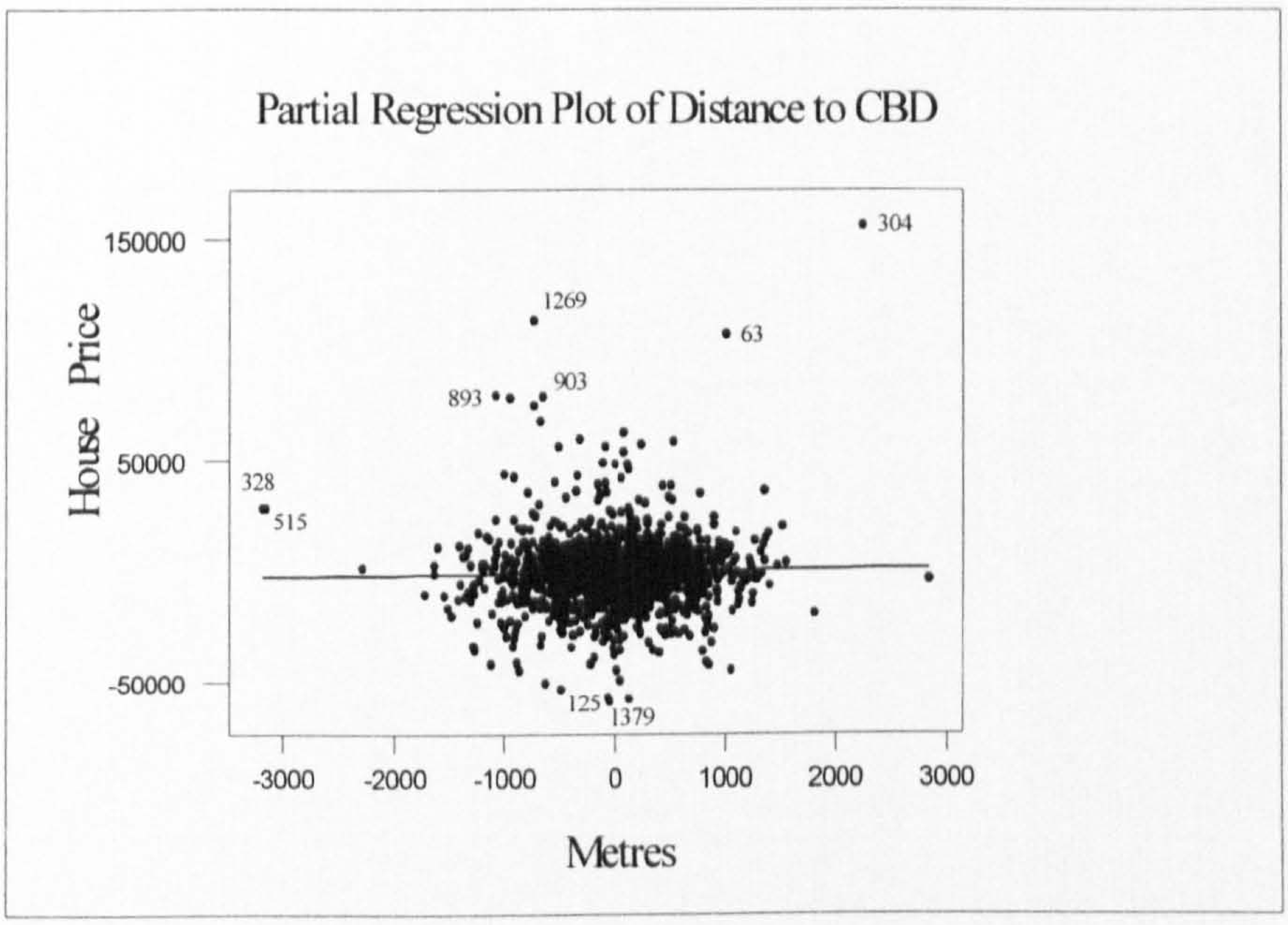
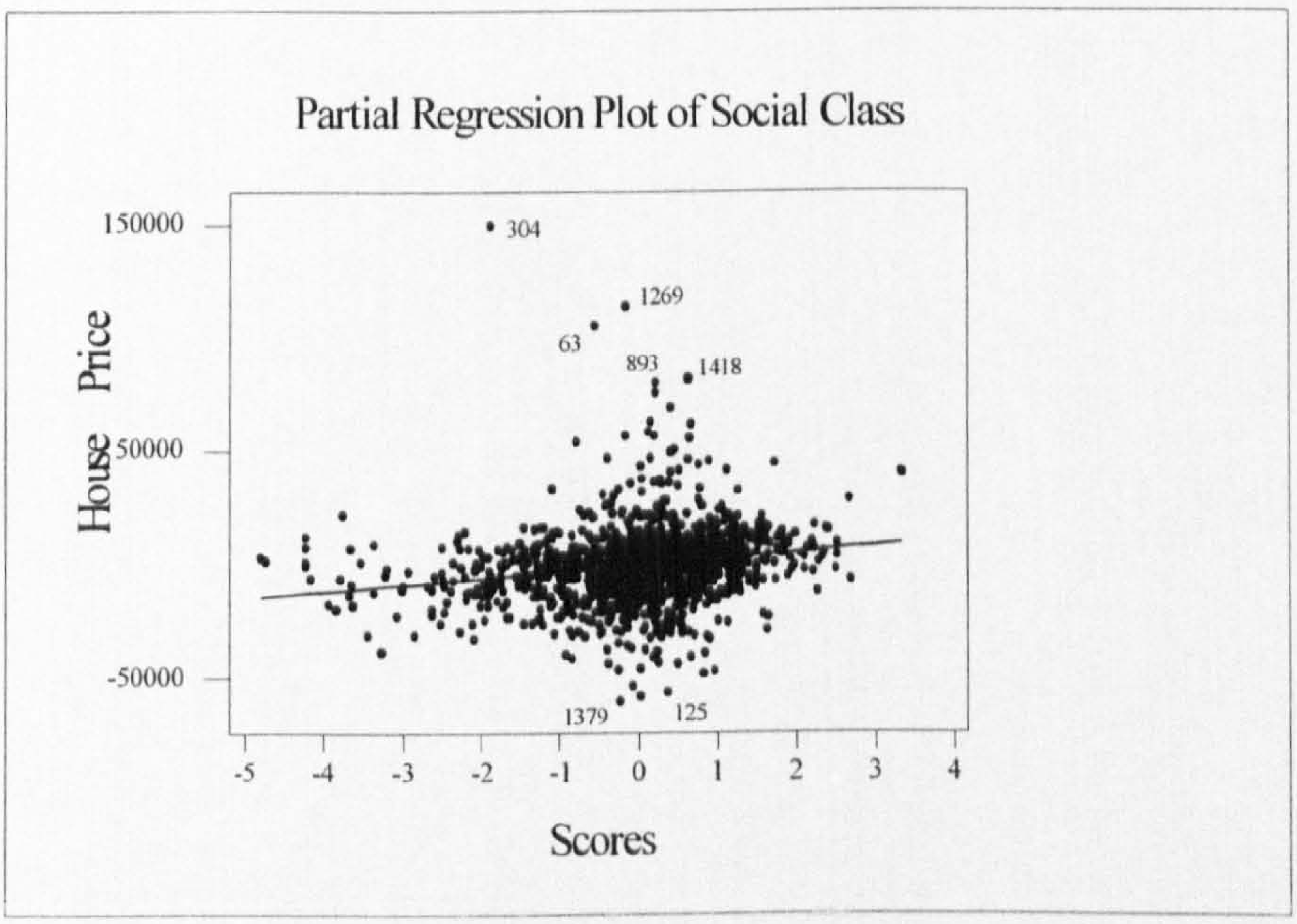


Figure 6.53 (cont.)





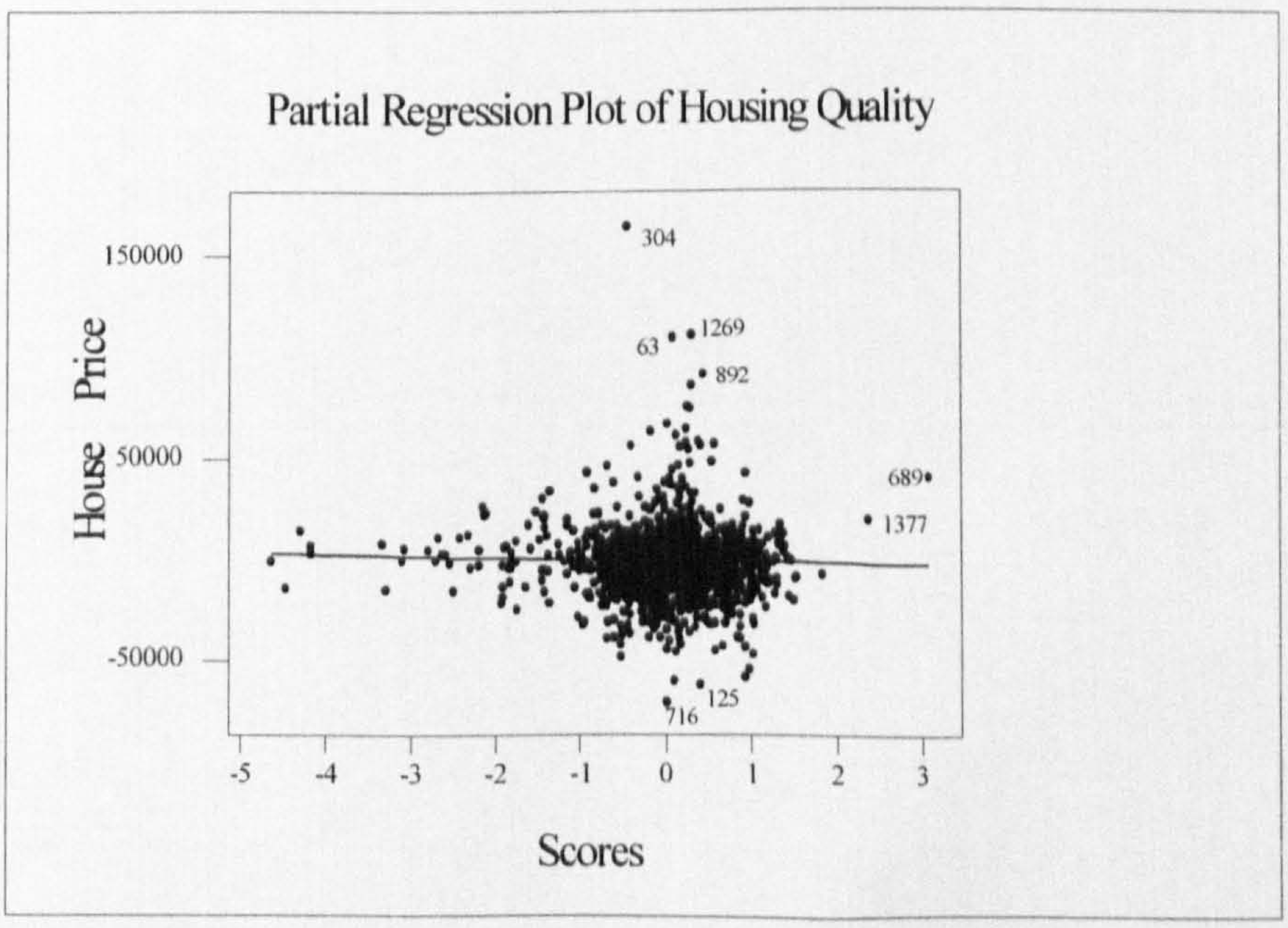
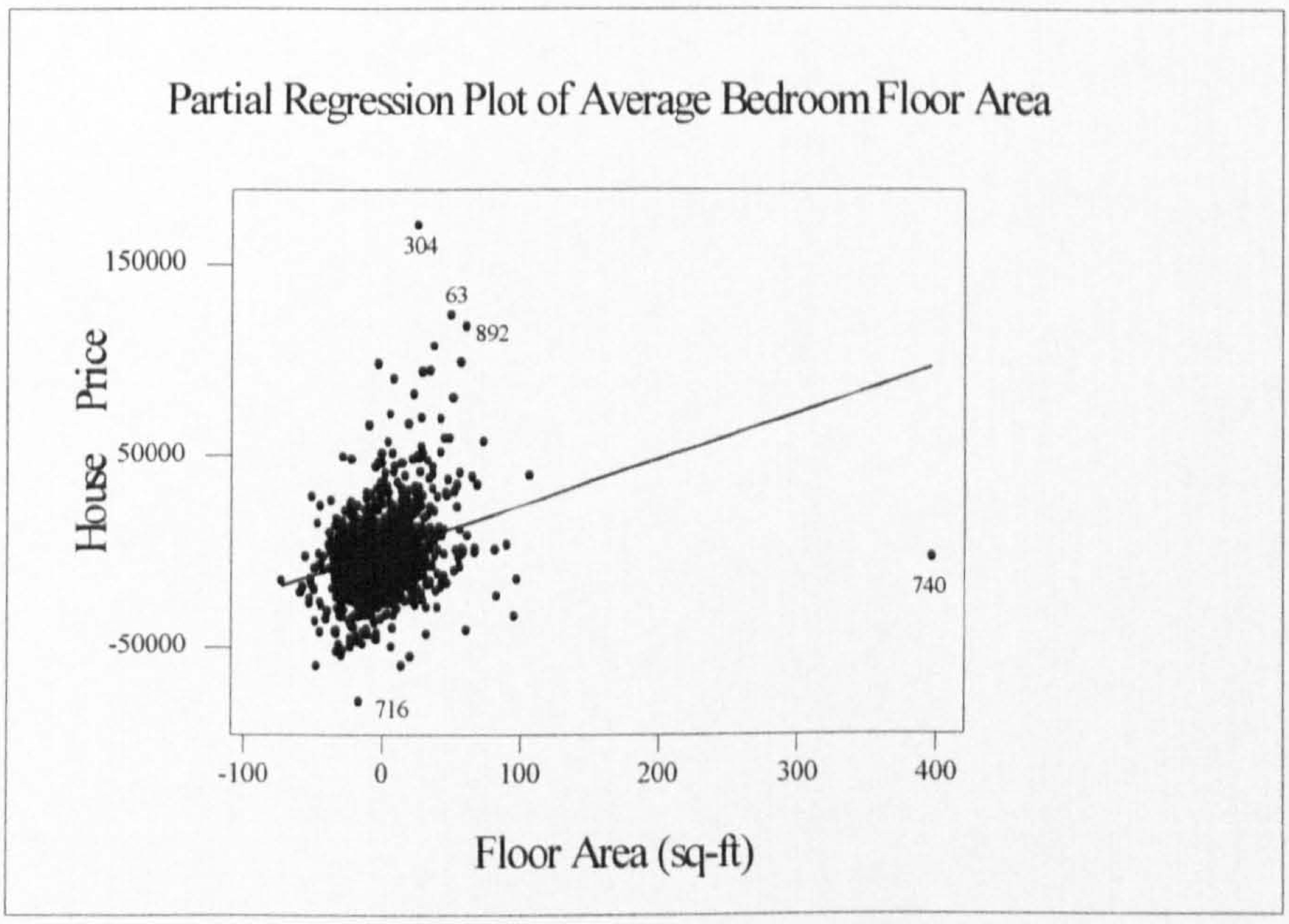
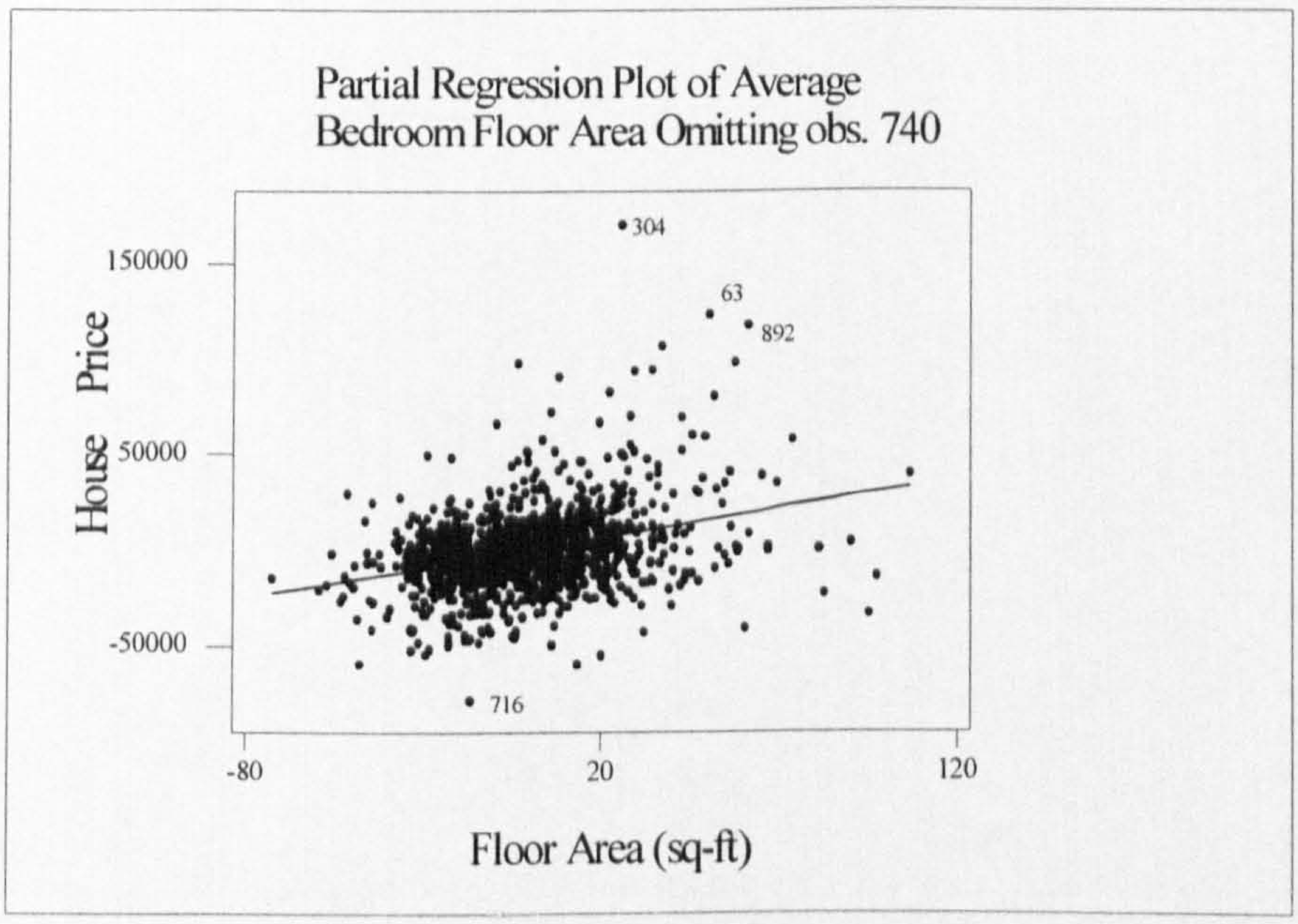
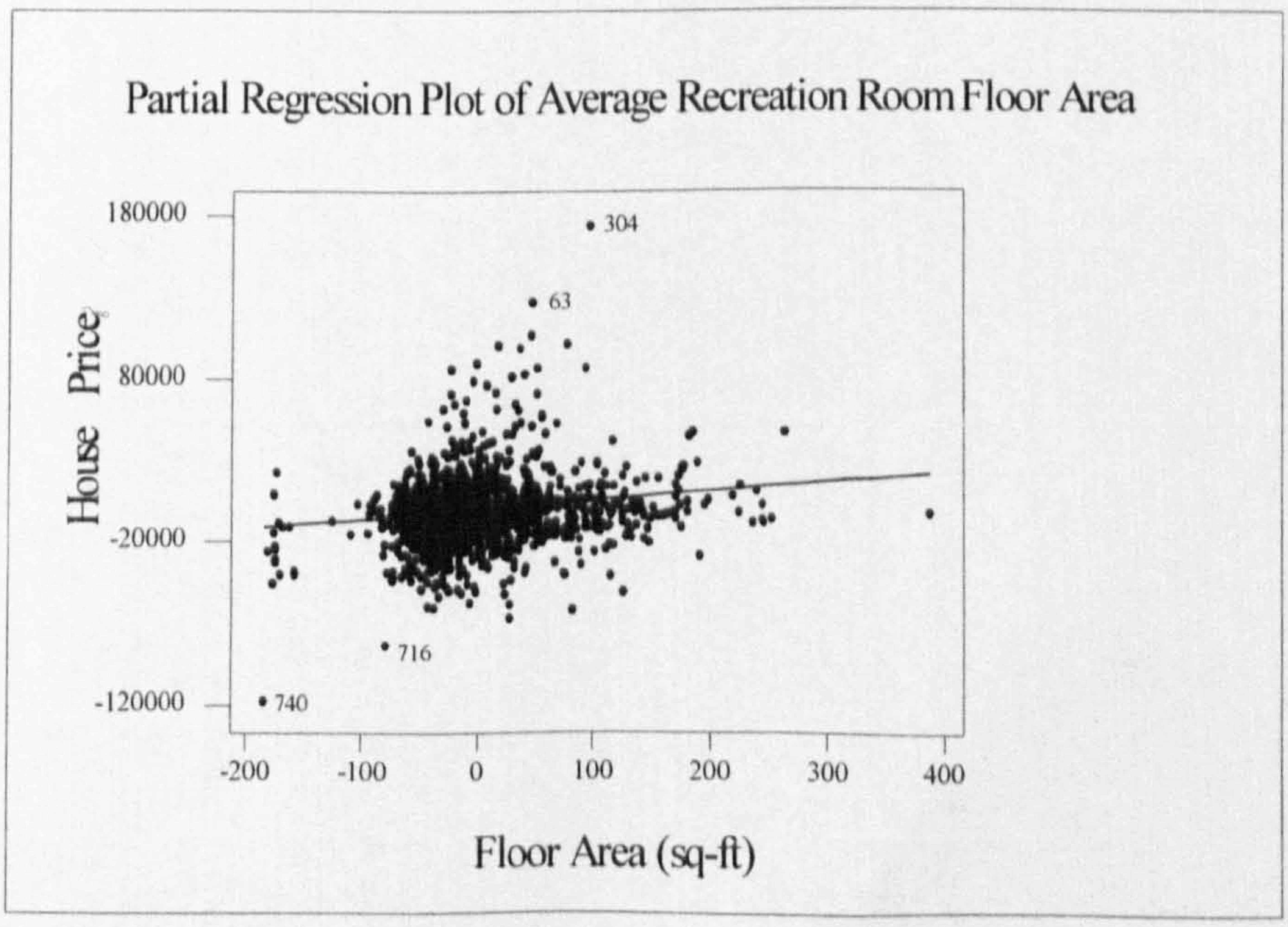


Figure 6.54: Selected Partial Regression Plots for Model 6.2







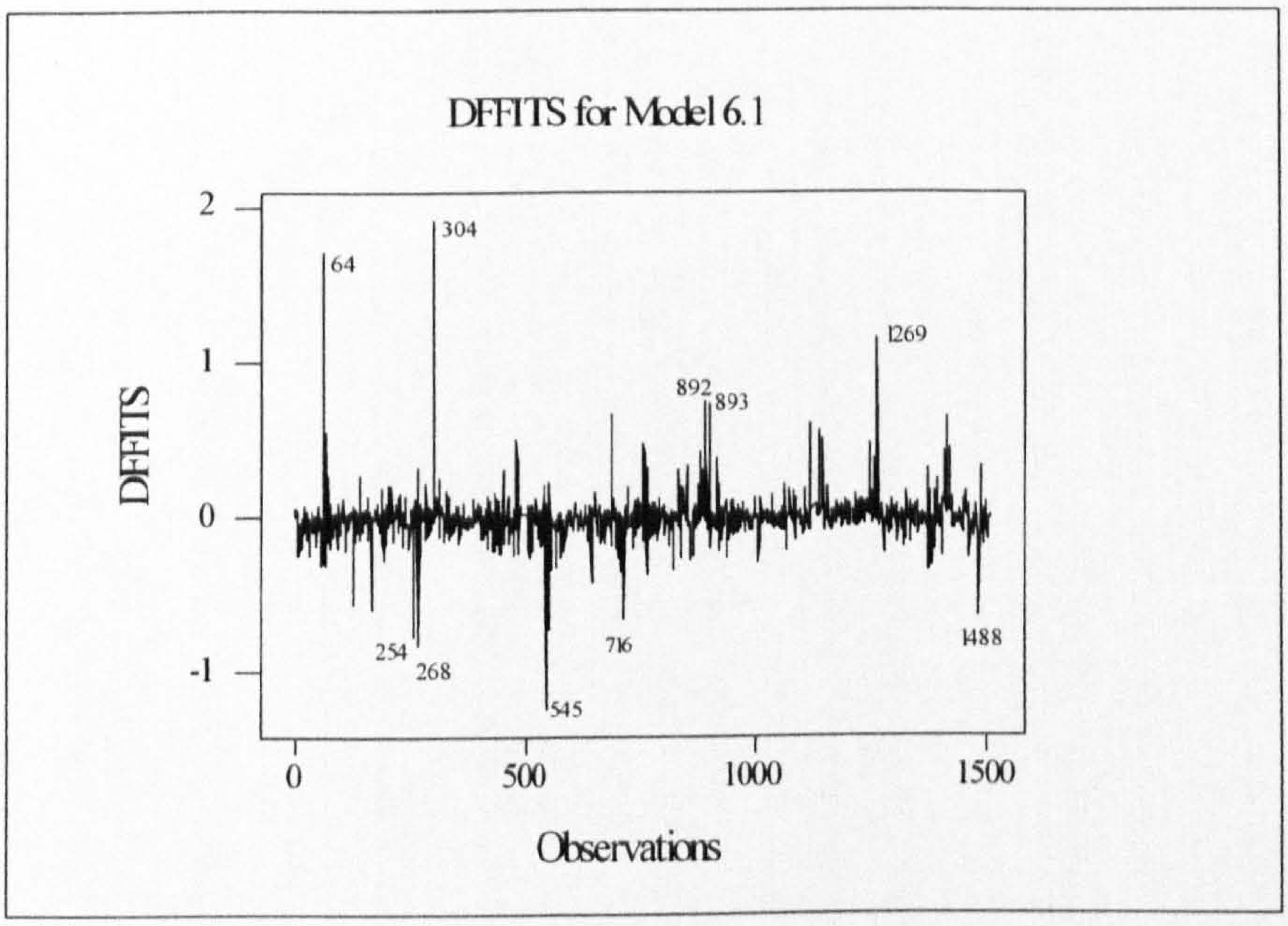
II. Partial Regression Plots

Previously, the relationship between the dependent and independent variables were assessed in terms of simple bi-variate plots. However, since the attributes within a hedonic model are, by nature, inter-related, it is expected that subsets of observations could be jointly influential or could offset each other's influence. To take this into account, partial regression plots were derived such that the relationship between the dependent and independent variable was established after the influence of all other independent variables had been statistically removed. These allow the partial or marginal contribution of an attribute in a hedonic model to be evaluated.

Figure 6.53 shows the partial regression plots of the continuously distributed attributes in model 6.1. It is evident that a number of outliers are present, with several data points lying a significant distance from the general clustering. Similar to the studentised residual plot, the majority of these outliers lie a significant distance above the regression line, suggesting that a number of properties have a greater price than predicted by the model. As was suggested by the box-plots, the plots show evidence of being positively skewed, with the majority of data points clustered around the mean. No plot shows evidence of non-linearities. This is particularly true for floor area. The simple bi-variate plots suggested that two separate linear functional relationships existed between house price and floor area. However, the partial regression plot suggests that, once all the other attributes have been taken into consideration, one continuous linear relationship exits. A similar conclusion can be reached for the social class and housing quality plot, whose previous simple bi-variate plots had implied a slight non-linear relationship.

Figure 6.54 shows a selection of the partial regression plots for Model 6.2. The majority were similar to those in Figure 6.53, with the exception of the room area variables. The plot of average bedroom size shows that the regression fit may be unduly influenced by observation 740. This may explain its insignificance in the model, and also the non-linear bi-variate plot in Figure 6.46. The plot also demonstrates how the data is clustered and lacks a great deal of variation. In contrast, the partial regression plot of average recreation room area appears to be more well behaved, and displays a definite linear functional form, as was suggested previously in Figure 6.46. Again, the outliers are similar to those in Model 6.1.

Figure 6.55: DFFITS for Model 6.1 and Model 6.2



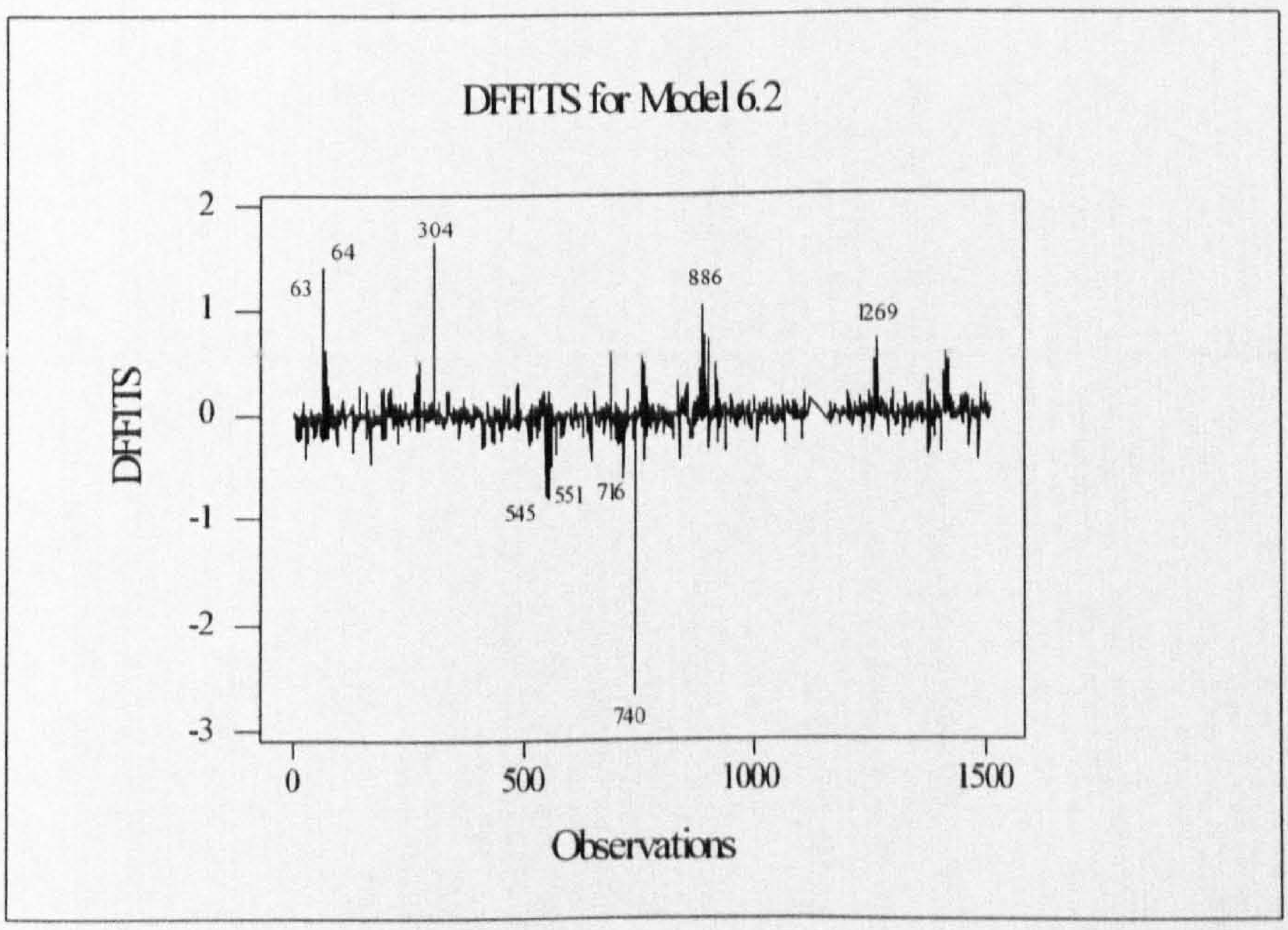
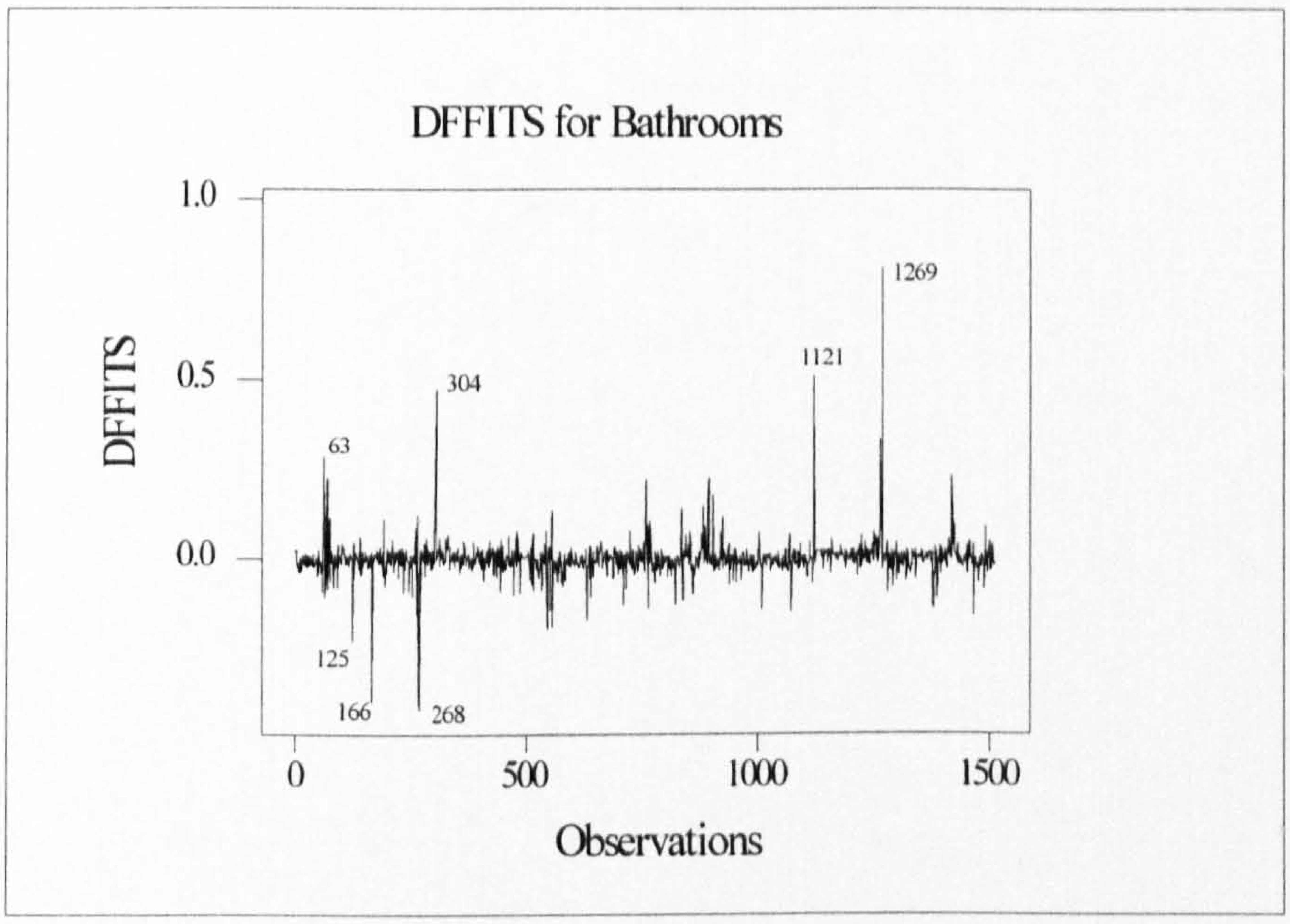
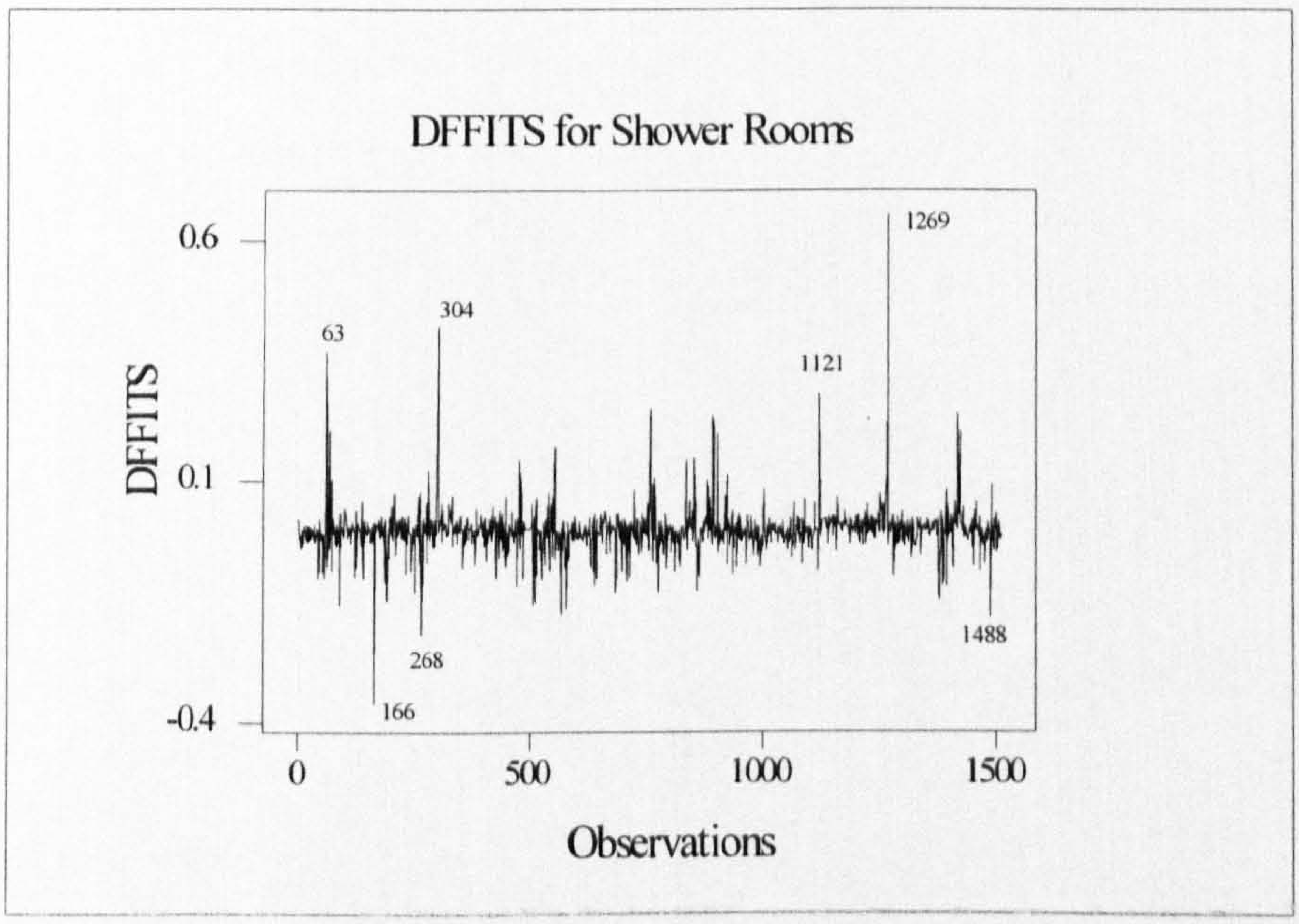


Figure 6.56: DFFITS for Model 6.1 Partial Plots

Figure 6.56 (cont.)





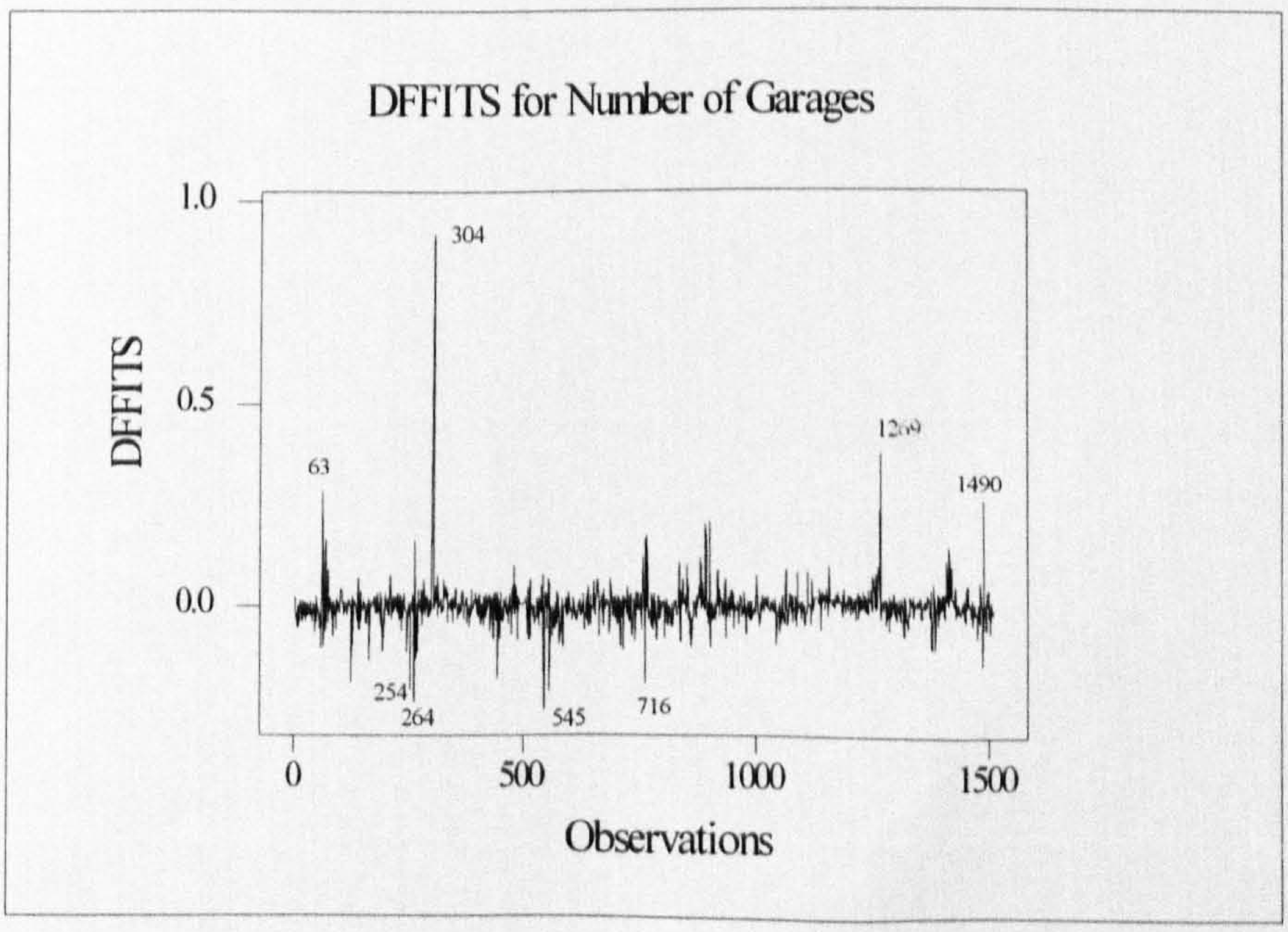


Figure 6.56 (cont.)

III. DFFITS

Joint diagnostic tests were applied to the partial regression residuals in an attempt to determine statistically those observations that had combined leverage and discrepancy effects. DFFITS were calculated for the full models and each of the continuously distributed variables, and were subsequently plotted against observation. A summary for the overall model and the partial regressions for Model 6.1 are presented in Figure 6.55 and 6.56. The cut-off for the DFFITS was calculated as an absolute value of 0.0745 (Belsley et al, 1980), and it can be clearly seen that several consistently large DFFITS occur in each plot. Similar results were discovered for Model 6.2, with the same observations indicated as significant. When these were analysed, it was discovered that twenty six observations were consistently more than twice the cut-off value across the majority of the variables in both models. Several of these observations had also previously been indicated as outliers in the partial regression plots.

Table 6.12 summarises the characteristics of these observations. This reveals that the properties were either large, expensive properties on the semi-rural fringe of Cardiff, or cheap, large properties in the Inner Area. Although a characteristic of large data sets such as this one is their inherent ability to absorb apparently anomalous data without changing the results significantly (Chatterjee and Price, 1977), it was decided to omit these observations from the study. The reasons for this lies in the fact that the majority of the properties lie within the upper quartile of both the house price and floor area distributions. An earlier examination of floor area had indicated possible heteroscedasticity with respect to these larger properties, and it is doubtful whether such large, semi-rural properties will be closely functionally related to the rest of the Cardiff property market. This is corroborated by an ANOVA test which proved that the mean floor area of the omitted observations in both models were significantly different at the 1% level (Model 6.1 F-statistic = 178.05, Model 6.2 F-statistic = 132.51; critical value $F_{1,1470} = 6.63$ at 1% significance level), suggesting that they may come from different distributions.

6.5.3.5 The Reduced Data Models

The two models were re-estimated using this reduced dataset and are summarised in Tables 6.13 and 6.14. The effect of removing the outlying observations are markedly different. In the total floor area model, half of the t-statistics decreased in value, although the majority of

Table 6.12
Summary of the Attribute of the Omitted Observations

Observation	House Price	Floor Area	Dwelling Type	Bedrooms	Bathrooms	Shower rooms	Garages	Garden Size	Neighbourhood
63	325000	1854.8	D	4	2	0	2	50m	Lisvane & St.Mellons
64	300000	1721.1	D	4	2	0	2	50m	Lisvane & St. Mellons
70	195000	1705.1	SD	5	1	0	0	50m	Llanshen
125	85000	1711.8	SD	4	2	0	0	5-50m	Plasnewydd
166	67950	1475.6	ET	5	2	1	0	0-5m	Plasnewydd
194	88950	1559.7	ET	5	1	1	1	0-5m	Roath
247	46950	1093	MT	4	1	0	1	0-5m	Riverside
254	199950	2843.96	SD	6	2	1	0	50m	Llandaff
264	179950	2272.2	D	5	1	1	0	50m	Llandaff
268	225000	2594.7	ET	9	4	1	2	0-5m	Riverside
304	395000	1909.9	D	4	2	1	3	50m	Lisvane & St. Mellons
545	139950	1367.3	B	4	2	0	0	50m	Cyncoed
557	56950	872.2	ET	3	2	0	2	0-5m	Splott
716	24950	598.7	B	2	1	0	1	50m	Rumney
740	228950	1588.6	D	5	1	0	1	50m	Whitchurch & Tongwylias
880	220000	1533.8	D	5	1	1	1	5-50m	Lisvane & St. Mellons
892	280000	1629.7	D	4	1	1	1	50m	Cyncoed
893	330000	2151.4	D	4	2	1	2	50m	Lisvane & St. Mellons
903	295000	1739.3	D	5	1	1	2	50m	Cyncoed
1121	245000	1328	D	5	3	1	2	50m	Rhiwbina
1264	125000	1295.8	MT	5	0	1	0	0-5m	Riverside
1266	180000	879.3	D	4	1	1	2	5-50m	Fairwater
1269	295000	1135.7	D	5	2	1	2	50m	Llandaff
1386	59950	687.5	D	3	1	0	1	50m	Lisvane & St. Mellons
1387	227500	1306.5	D	4	1	0	1	50m	Heath
1488	72000	1264.7	FCB	3	1	1	1	5-50m	Llandaff

Table 6.13

Model 6.3: Reduced Dataset Total Floor Area

Predictor	Coeff	St.Error	T-stat	VIF
Constant	53790	97.00	556.40	
Floor Area	31.30	2.19	14.30	5.0
ET	-130	1300	-0.10	4.4
SD	4281	2183	1.96	6.1
D	16451	1920	8.57	4.8
FPB	1887	2074	0.91	6.8
FCB	1109	1879	0.59	2.8
M	-1057	2158	-0.49	2.3
B	18839	2272	8.29	2.3
EL	-801	1742	-0.46	2.4
ML	-2750	1528	-1.80	3.1
Beds	807	829	0.97	4.5
Recs	888	747	1.19	2.2
Baths	8692	1439	6.04	1.4
Showers	5920	1059	5.59	1.5
Full CH	4314	1319	3.27	2.8
Part CH	-56	1882	-0.03	1.4
Gas	1326	1028	1.29	2.5
Garage	3245	683	4.75	1.8
ORP	3640	769	4.73	1.8
Age: New	3679	2343	1.57	1.2
Post 1964	-282	1280	-0.22	5.1
1918-64	2285	1270	1.80	3.4
Gdn: None	-3121	1530	-2.04	7.4
Gdn: 5-50m	3091	997	3.10	6.4
Gdn: >50m	6440	1563	4.12	4.3
Cons	1957	1262	1.55	1.0
Needs Mods	-6191	1326	-4.67	1.2
Swm Pool	3175	6225	0.51	1.1
Dist CBD	-2.08	0.24	-8.78	3.0
Social	4939	262	18.83	2.7
H.Qual	-1392	389	-3.57	1.7
LA > 50%	3261	1614	2.02	2.0
s	14206		R-sq(adj)	84.1

Table 6.14

Model 6.4: Reduced Dataset Average Floor Area

Predictor	Coeff	St.Error	T-stat	VIF
Constant	51330	102	503.50	
Ave Bed Floor	70.06	35.56	1.97	8.3
Ave Rec Floor	46.05	5.49	8.39	4.8
Ave Kit Floor	37.76	7.22	5.23	1.2
ET	-712	1395	-0.51	4.2
SD	3649	1317	2.77	6.9
D	15730	1976	7.96	5.7
FPB	2791	2215	1.26	6.5
FCB	660	2062	0.32	2.2
M	-375	2205	-0.17	2.5
B	17162	2421	7.09	2.2
EL	-497	1841	-0.27	2.5
ML	-1198	1619	-0.74	3.2
Beds	7399	548	13.50	2.7
Recs	7632	6255	1.22	1.4
Baths	9439	1470	6.42	1.3
Showers	5636	1048	5.38	1.3
Full CH	5477	1401	3.91	2.8
Part CH	1253	2054	0.61	1.5
Gas	-155	1032	-0.15	2.5
Garage	2793	702	3.98	1.8
ORP	3732	810	4.61	1.9
Age:New	6750	2627	2.57	1.3
Post1964	-1402	1263	-1.11	5.0
1918-64	918	1293	0.71	3.4
Gdn: None	-2561	470	-5.45	6.6
Gdn: <5m	2960	1039	2.85	5.8
Gdn: 5-50m	6308	1573	4.01	3.7
Cons	2531	2556	0.99	1.0
Needs Mods	-5397	1447	-3.73	1.2
Swm Pool	6078	4749	1.28	1.1
Dist CBD	-1.74	0.24	-7.22	3.0
Social	4746	270	17.60	2.7
H.Qual	-1327	423	-3.14	1.7
LA > 50%	2157	849	2.54	1.9
s	20422		R-sq(adj)	75.7

these were insignificant in Model 6.1. The locational attributes became of increased significance, whilst the garden variables and some of the property type variables became statistically less important. However, the number of bedrooms variable, which previously had a counter-intuitive relationship with house price, changed sign. This confirms the suspicions that the number of bedrooms variable was influenced by outliers, although this and the number of recreation rooms remained insignificant. The overall estimated standard error decreased, suggesting that the model fits the data better.

In comparison, the majority of t-statistics in the average floor area model increased, including average bedroom floor area, which became significant. However, the number of bedrooms still remained the preferred explanatory bedroom variable of house price variation. The degree of multicollinearity in the model also declined, although the overall estimated standard error increased, reducing the model's fit. Inspection of the partial residual plots from these models showed no obvious violations of the assumptions. The next stage was to systematically drop the insignificant variables in each of the models in a stepwise fashion, the order being determined by the magnitude of the t-values. The models were re-estimated after each variable was removed, with the final models summarised in Tables 6.15 and 6.16.

6.5.3.6 The Final Models

All the variables in the final total floor area model (Model 6.5) are well behaved with respect to their signs, and are significant at the 5% level (1.96), with the majority also significant at the 1% level (2.58). The VIFs suggest that multicollinearity is not a problem. This model explains more of the variation in house price than the previous model, and the t-statistics are also higher in magnitude. The number of recreation rooms has been dropped from the model, as have the age variables and several house type categories through lack of significance. New partial regression plots and DFFITS were estimated, and these showed no evidence of strongly influential outliers. Figure 6.57 shows the partial regression plot of floor area, and Figure 6.58 the bi-variate LOWESS plots for floor area, distance to the CBD, social class and housing quality. The floor area plots confirms the linear, and possible heteroscedastic relationship with house price, whilst the break in the slope implied in Figure 6.43 is no longer evident. This suggests that this was an artefact of the multivariate nature of the data, and has subsequently been accounted for by the inclusion of other variables. The LOWESS plots for distance to the city centre now reveals the hypothesized negative, linear

Table 6.15

Model 6.5: Final Total Floor Area

Predictor	Coeff	St.Error	T-stat	VIF
Constant	44882	69	649.21	
Floor Area	35.10	1.12	31.73	2.0
SD	3184	782	4.07	2.0
D	16448	1257	13.08	2.3
B	15977	1607	9.94	1.3
Baths	6478	1186	5.46	1.4
Showers	4949	880	5.62	1.4
Full CH	4568	773	6.32	1.2
Garage	3146	566	5.56	1.7
ORP	2825	596	4.74	1.6
Gdn:None	-2926	758	-3.86	6.1
Gdn:5-50m	2931	755	3.88	5.7
Gdn:>50m	5519	1237	4.46	4.2
Needs Mods	-4628	11123	-4.16	1.1
Dist CBD	-1.80	0.17	-10.57	2.5
Social	4077	210	19.42	2.6
H.Qual	-1388	255	-5.44	1.4
LA > 50%	3122	1323	2.36	1.9
s	13965		R-sq(adj)	83.4%

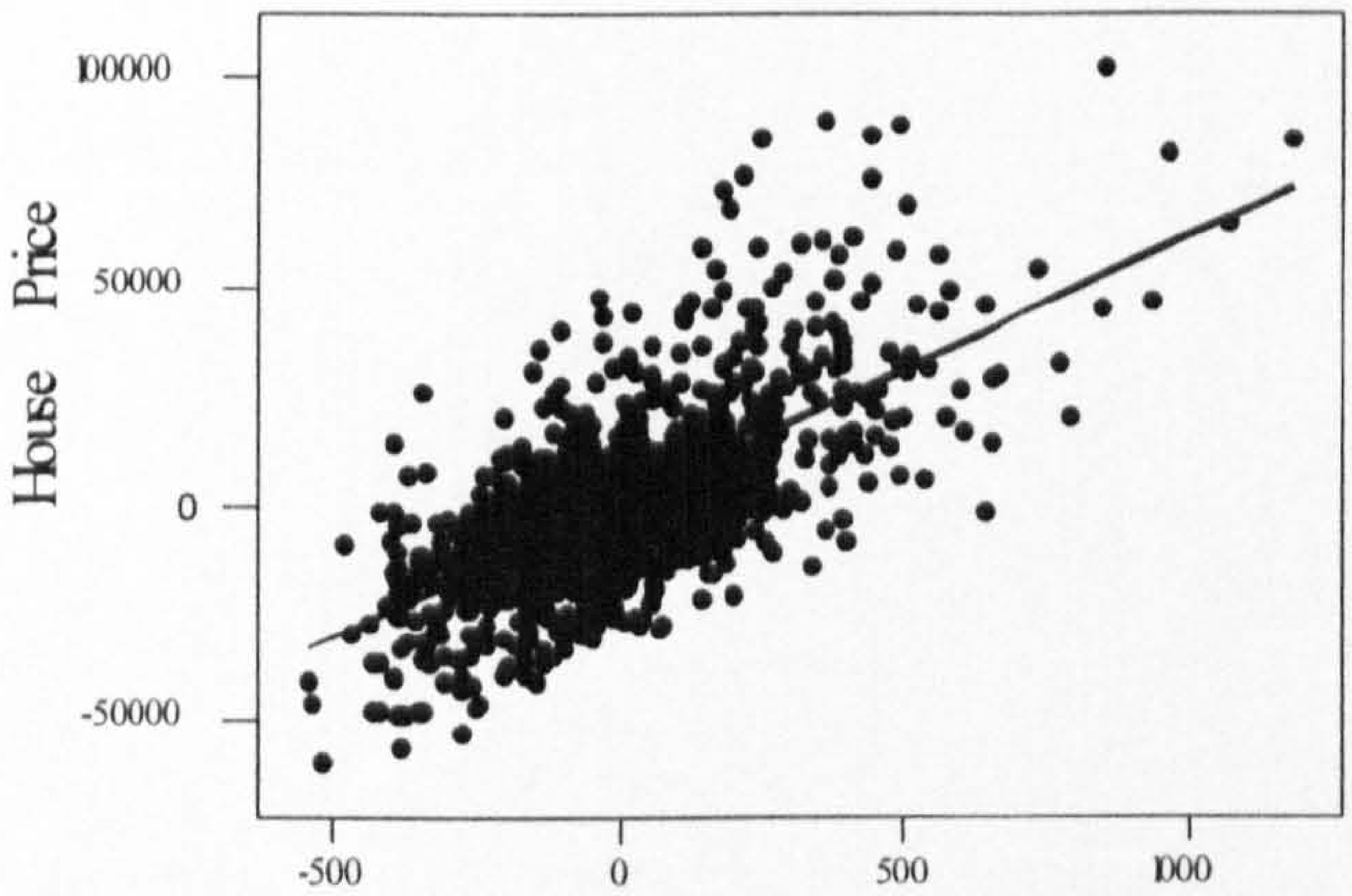
Table 6.16

Model 6.6: Final Average Floor Area

Predictor	Coeff	St.Error	T-stat	VIF
Constant	46770	43.87	553.66	
Ave Bed Floor	74.07	36.49	2.03	7.9
Ave Rec Floor	14.78	4.68	3.16	4.2
Ave Kit Floor	21.74	6.86	3.17	1.3
SD	4215	910	4.63	2.6
D	17360	1469	11.82	2.3
B	19752	2062	9.58	3.1
Beds	9448	853	11.07	1.4
Baths	10782	1434	7.52	1.4
Showers	6838	1024	6.68	1.4
Full CH	4424	873	5.07	1.9
Garage	3119	681	4.58	1.9
Gdn:None	-2643	479	-5.52	6.7
Gdn:5-5m	2423	910.9	2.66	4.2
Gdn:>50m	8679	1454	5.97	3.8
Dist CBD	-2.15	0.21	-10.18	3.0
Social	4781	258	18.54	2.7
H.Qual	-1185	404	-2.93	1.8
LA > 50%	2238	986	2.27	1.5
s	20568		R-sq (adj)	75.3%

Figure 6.57

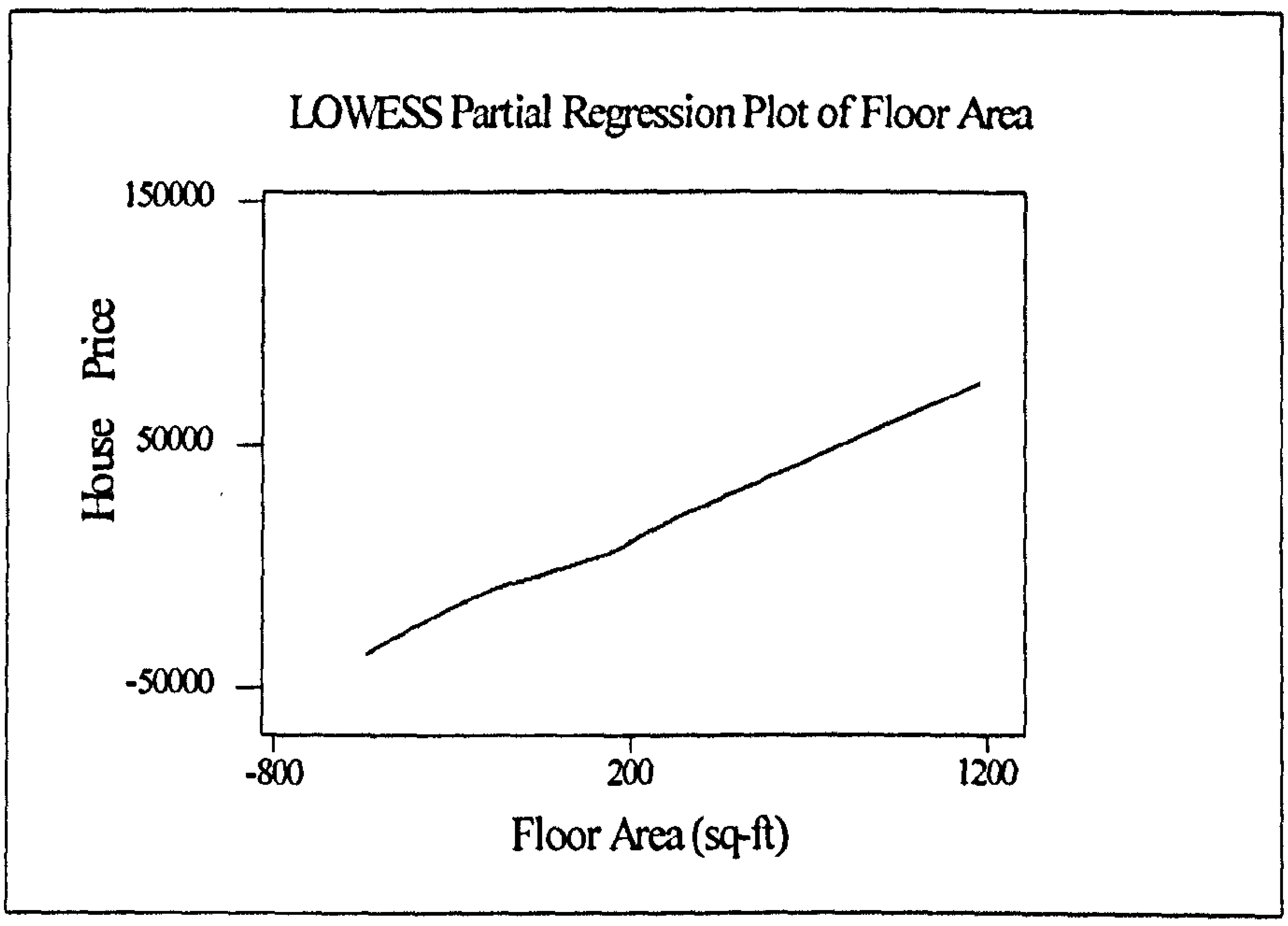
Partial Regression Plot of Floor Area for Model 6.5



relationship with house price, confirming the suspicions that the positive trend in Figure 6.48 was caused by uncontrolled for effects. A similar conclusion can be reached for the relationship between social class and house price, which was previously suggested to be non-linear for areas associated with higher social class. However, the housing quality variable continues to portray a non-linearity with house price, although this is very slight. The studentised residuals and DFFITS for the full model (Figure 6.59) continued to display several potential outliers, but these were fewer than in Model 6.1, and were deemed to be less problematic given the outcome of the partial plots and the size of the sample.

The final average floor area model (Model 6.6) has similar variables to the total floor area model, and the t-values are stronger on the whole. However, this is a consequence of the housing attributes, such as number of bedrooms, compensating for the poor performance of the floor area variables compared to Model 6.5. The variable that measures whether the property is in need of modernisation has also failed to be significant. The VIF indicates that some multicollinearity is present with respect to the average bedroom variable, but generally this appears not to be problematic. Moreover, the R-squared statistic indicates that the model only explains three quarters of the variation in house price, compared to over eighty percent in the full floor area model. The LOWESS partial regression plots for average floor area of the habitable rooms (Figure 6.60) suggest that the linear functional forms used to estimate the relationship with house price was the most suitable. The

Figure 6.58: Selected LOWESS Plots for Partial Derivatives of Model 6.5



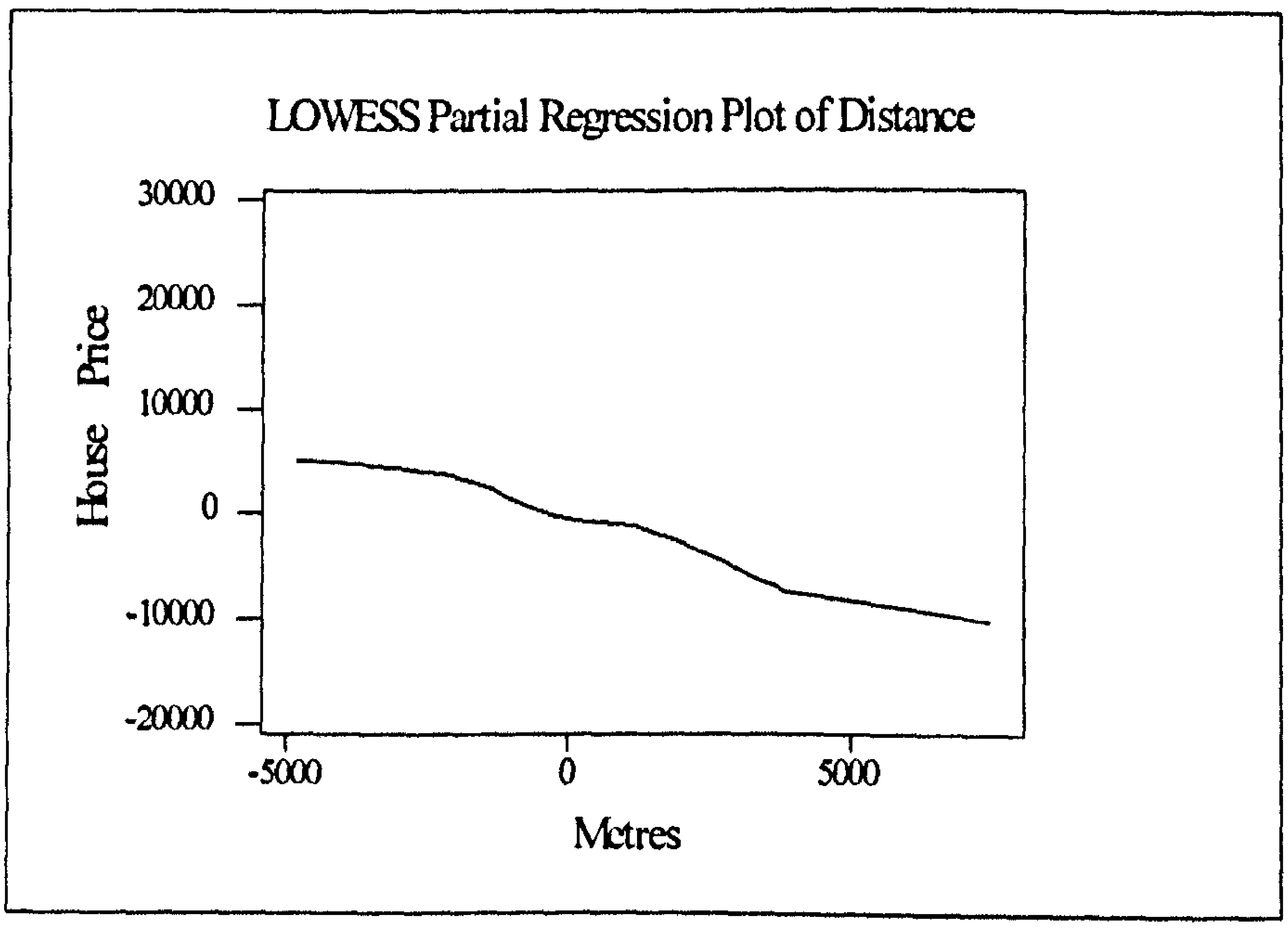
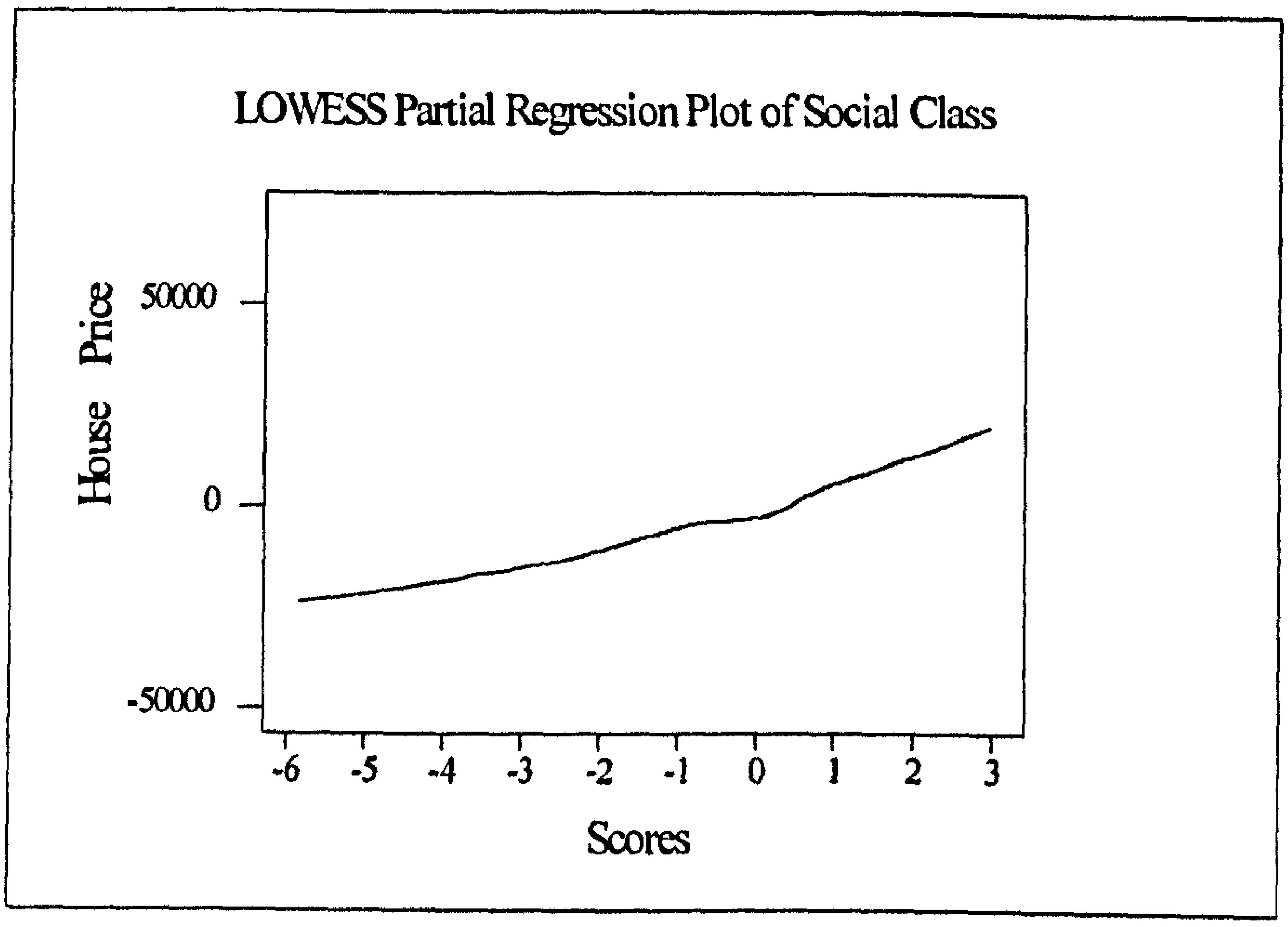


Figure 6.58 (cont.)



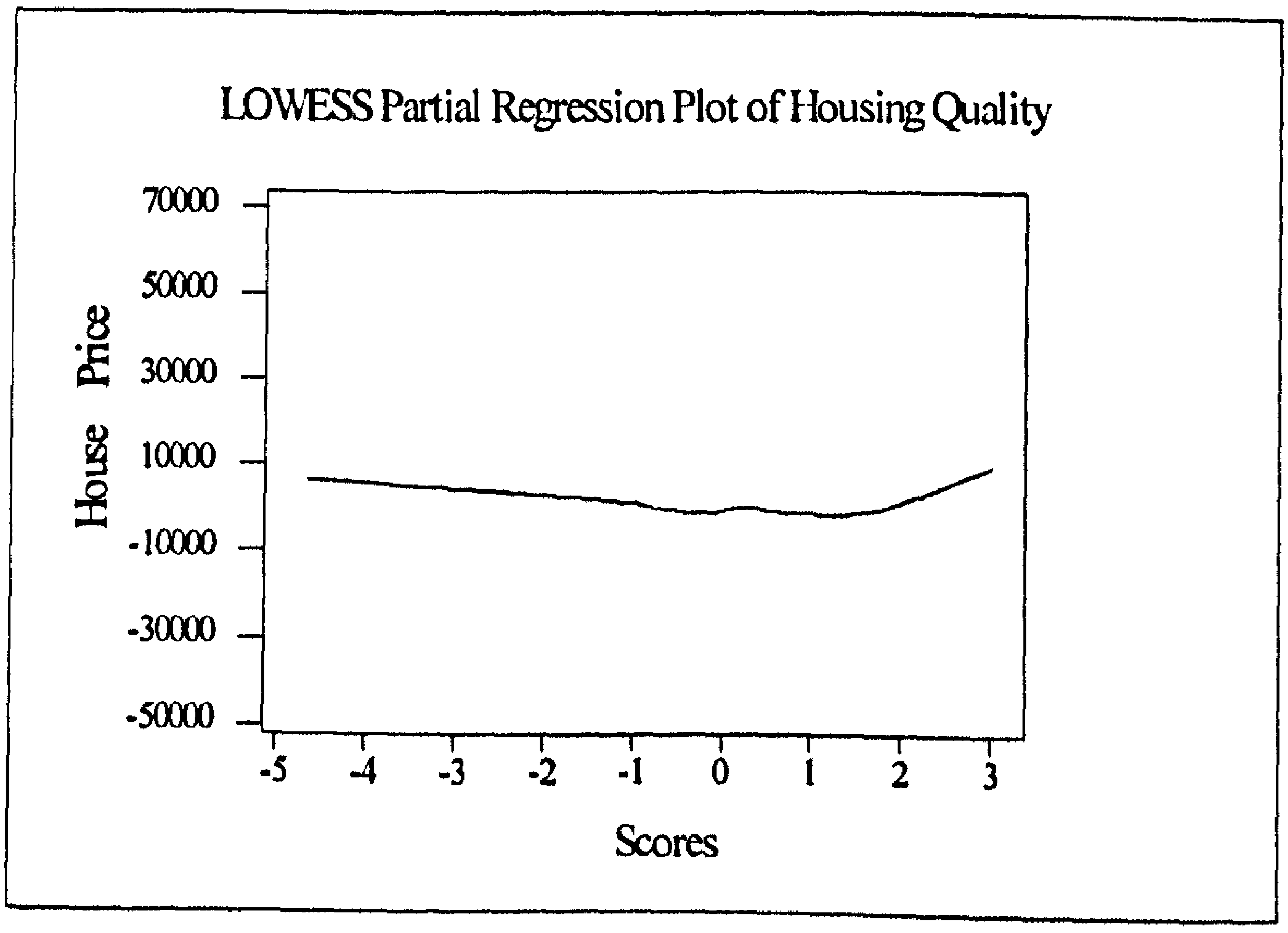
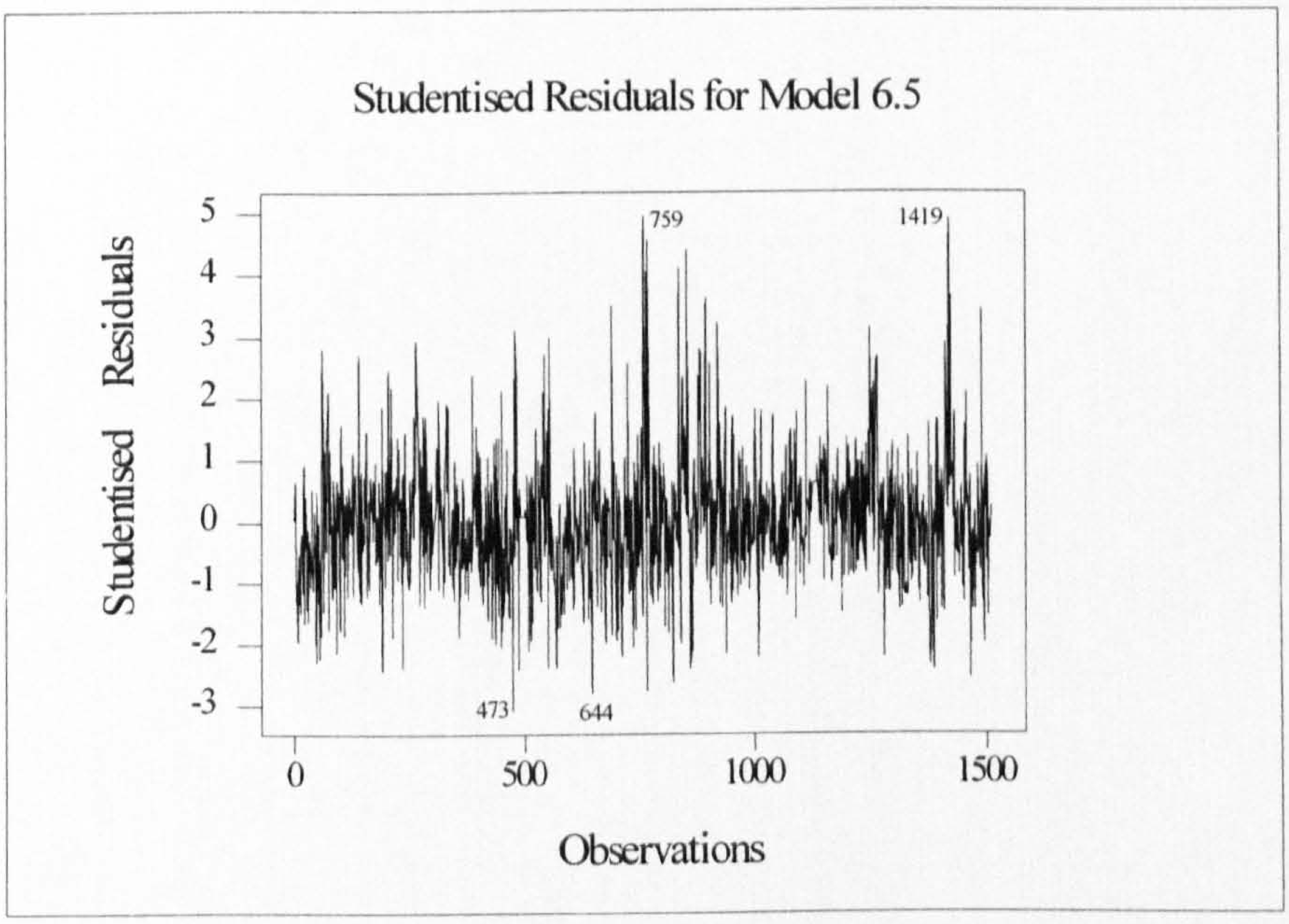


Figure 6.59: Model 6.5 Residual Diagnostics



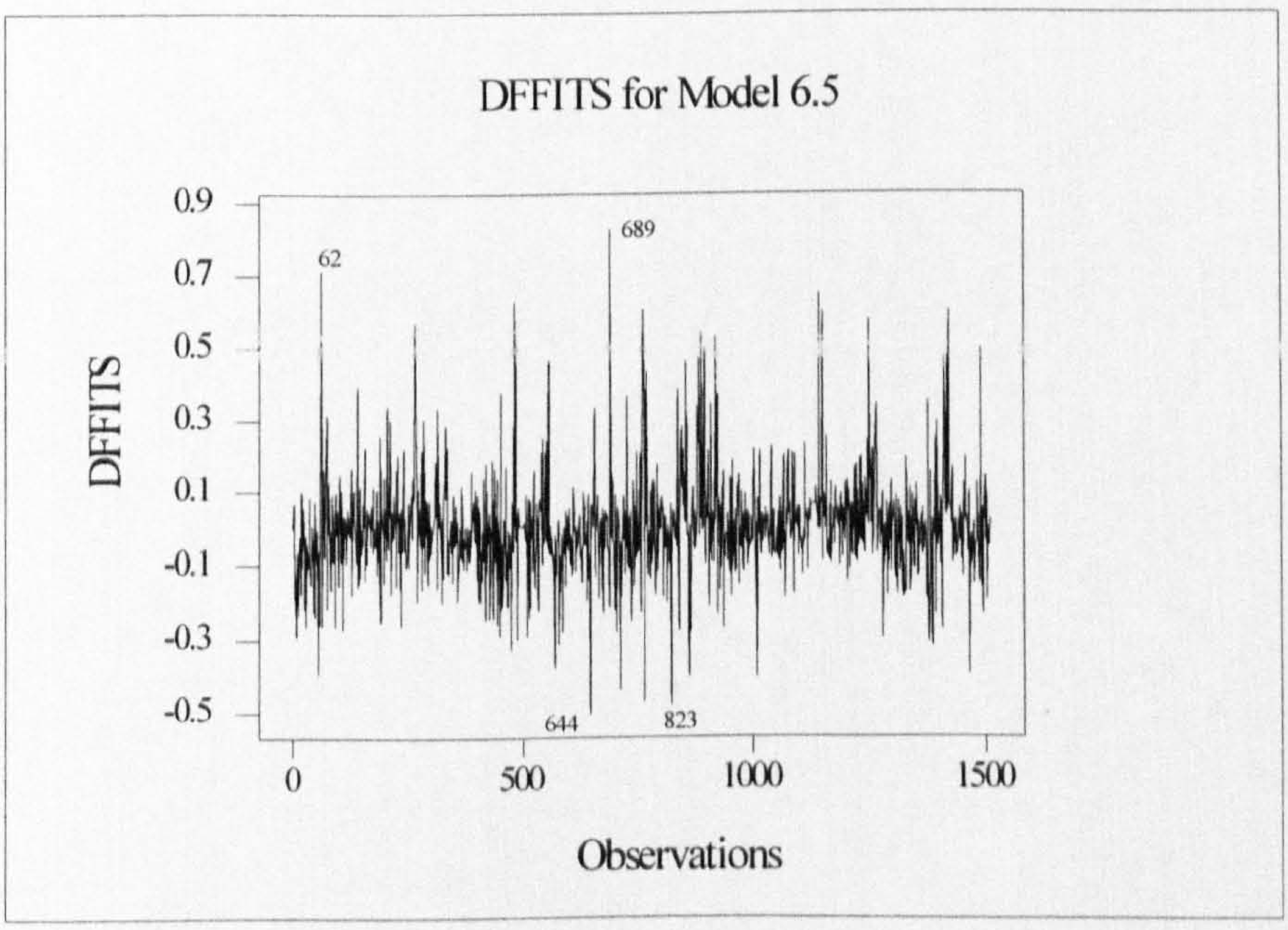
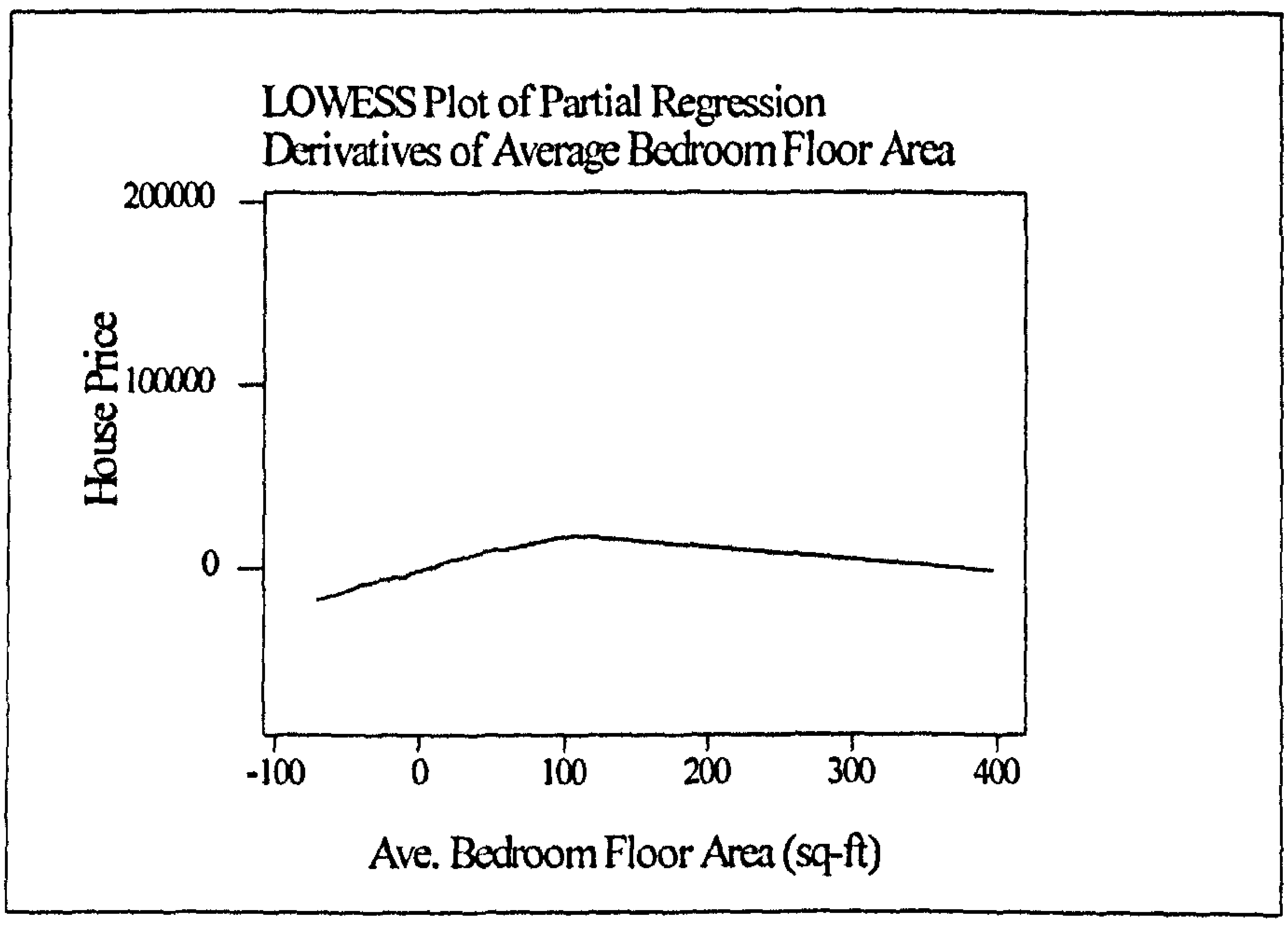
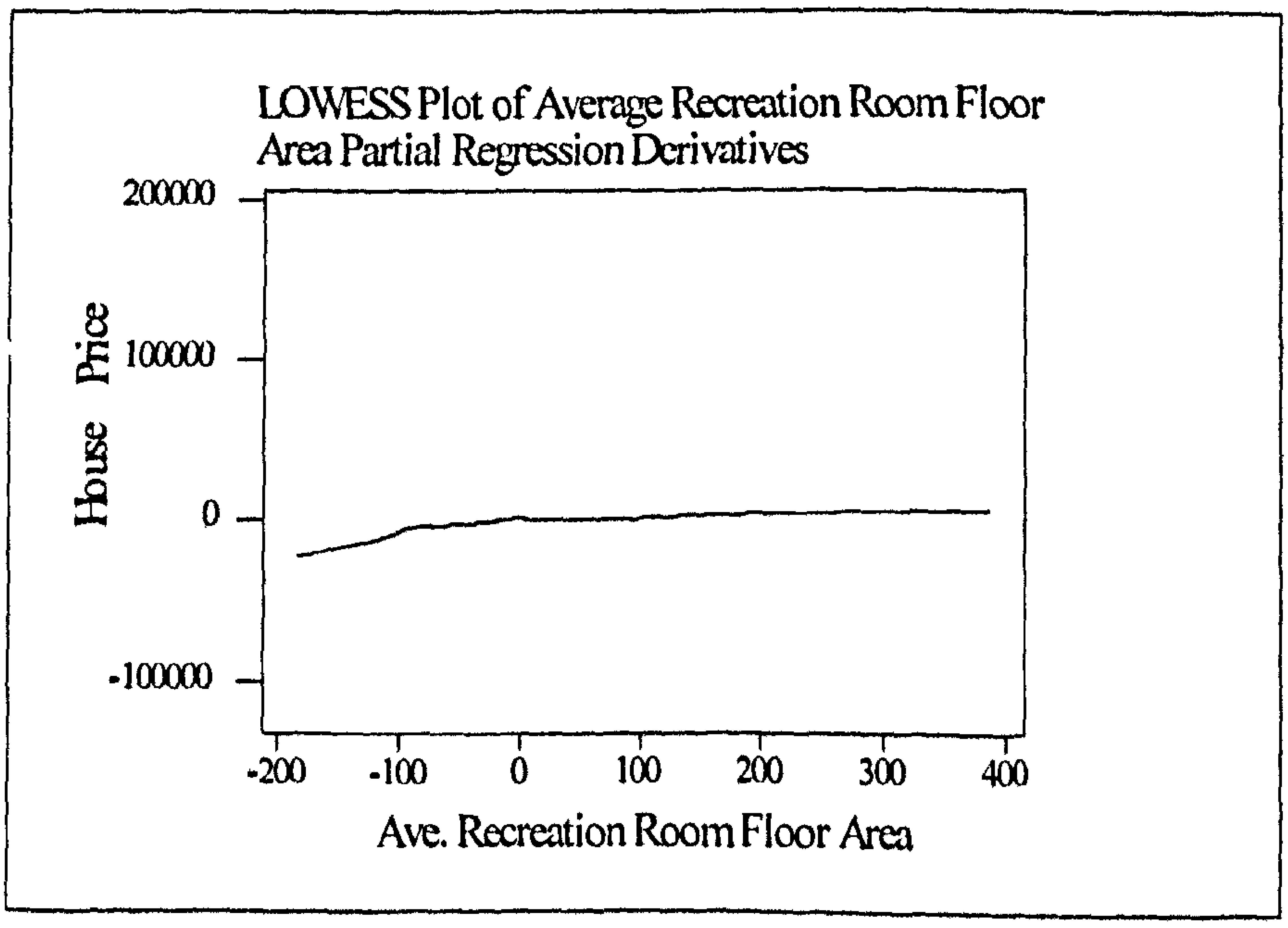


Figure 6.60: Selected LOWESS Plots for Partial Derivatives of Model 6.6





studentised residuals are very similar to Model 6.5, although the DFFITS for the full model indicate that the persisting outliers may have a stronger influence - see Figure 6.61

6.5.3.7 Error Term Diagnostics

The above analysis has described the process of building two hedonic models in an effort to best explain the structure of the data for the entire Cardiff housing market. Specifically, it has concentrated upon standard statistical and graphical diagnostical tests to check for violations of the first three assumption of the OLS regression model. This was necessary before the assumptions of independent, identically distributed errors could be checked since violations of the former can be exhibited in the error term as violations of the latter. Therefore, the remaining part of the section will test for the presence of heteroscedasticity and autocorrelation in the final two models.

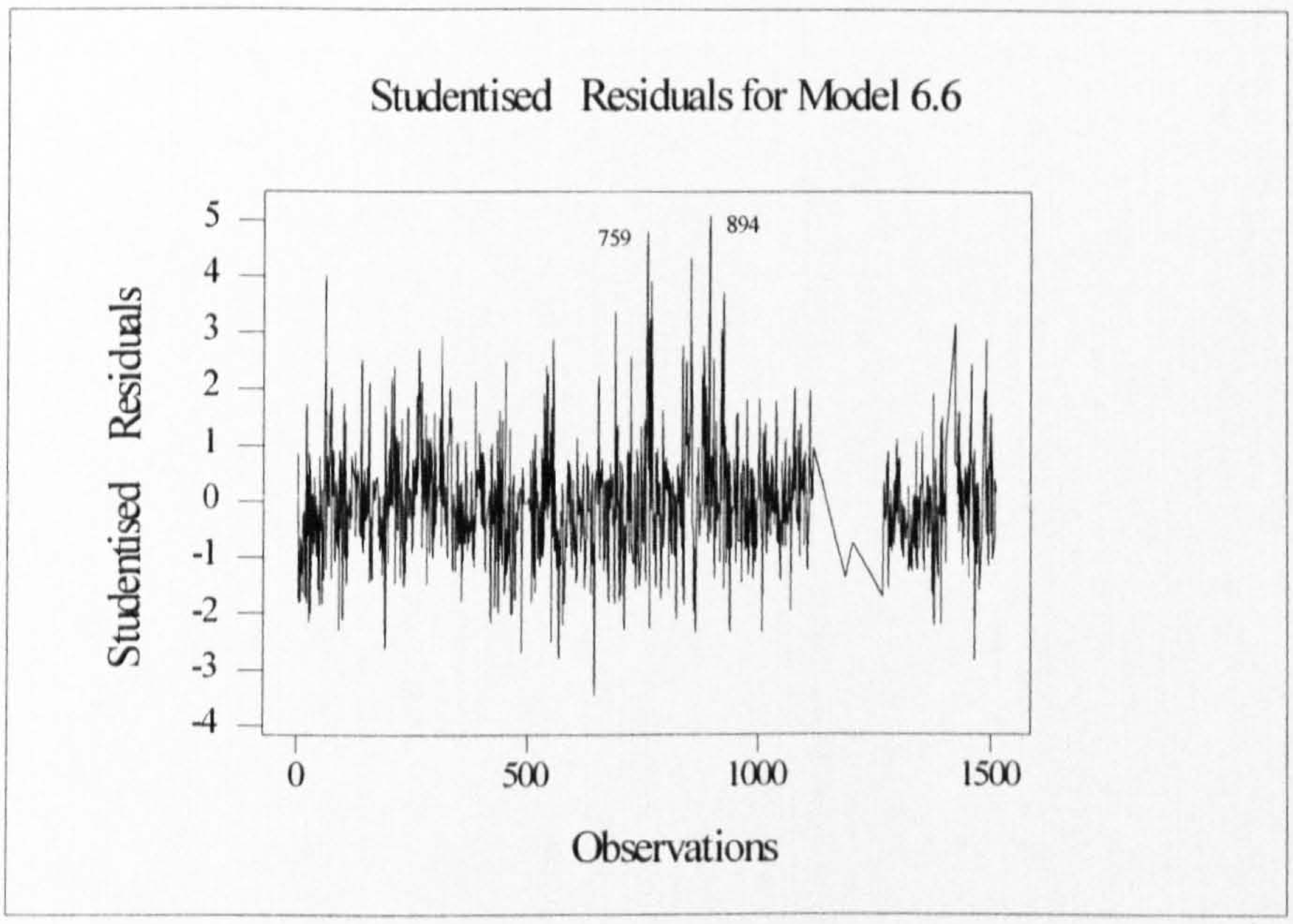
I. Heteroscedasticity

Heteroscedasticity can be a major problem in hedonic models. It will make the model highly inefficient, and cause the OLS estimation to place more weight on the observations with large error variances than on those with smaller error variances. The result will be biased, inefficient estimated variances of the estimated parameters which can cause statistical tests and confidence intervals to be incorrect . *Chapter Two* discussed several reasons why heteroscedasticity can be expected in hedonic models. Firstly, housing market segmentation may cause spatial parameter drift and therefore heteroscedasticity in the error term. Secondly, since the housing bundles cannot be untied and repackaged as desired, the attributes of different housing bundles may have different implicit prices. Hence, structural parameter drift may also be expected if the housing bundles are contextually different. Thirdly, a sample may be drawn from a distribution that has non-constant variance, and therefore a variable may be inherently heteroscedastic. For instance, the marginal price of floor area may be more variable in larger houses than smaller ones, due to greatly variability in supply and demand dynamics in the upper end the of the housing market.

The presence of heteroscedasticity in the error term can be checked by plotting the residuals of the hedonic model against the predicted house prices. A non-random plot is indicative of heteroscedasticity, particularly if the residuals increase with the predicted values. Figures 6.62 are plots for Model 6.5 and Model 6.6 respectively. Although clustering of the errors is

Figure 6.61:

Model 6.6 Residual Diagnostics



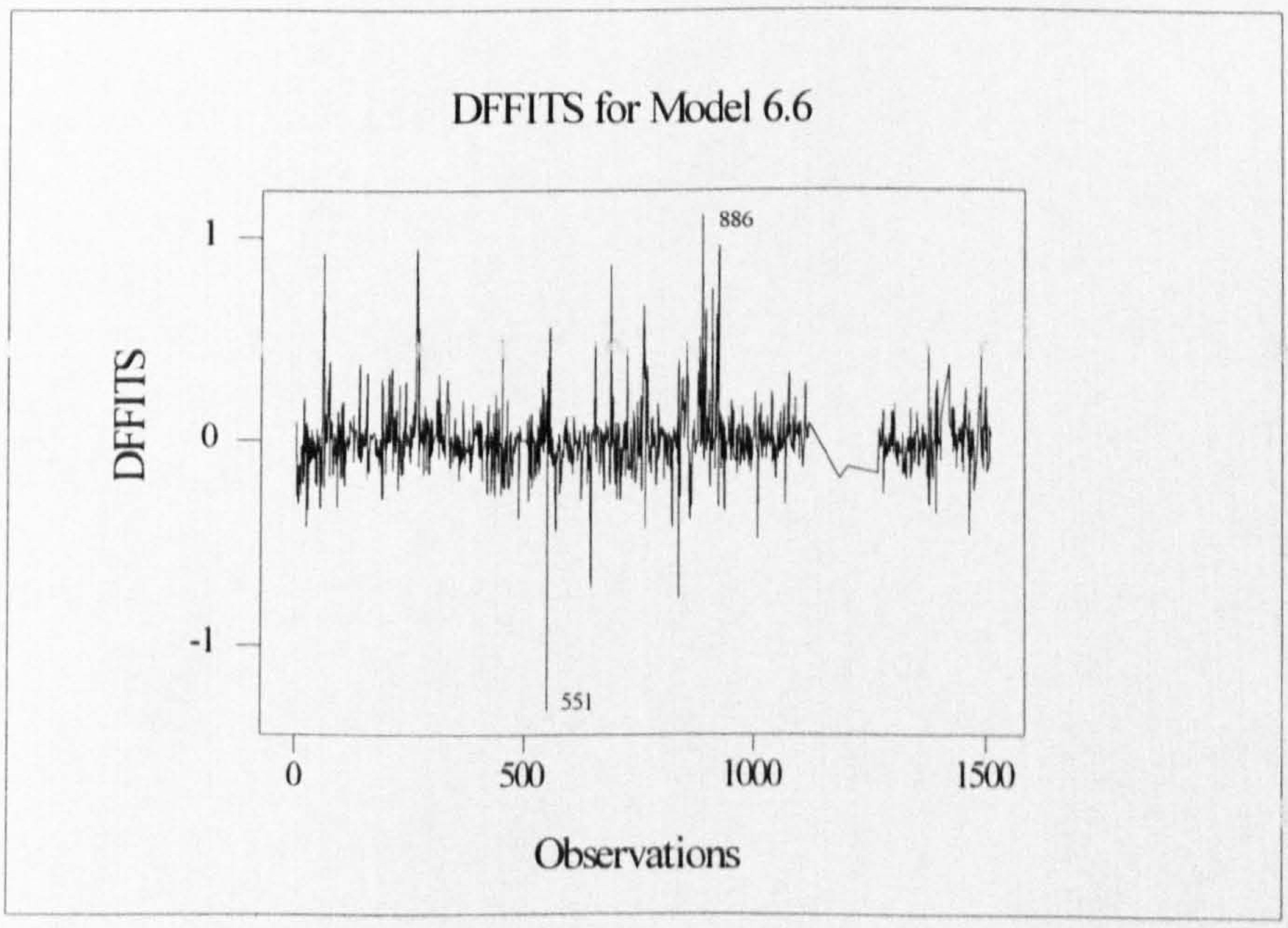
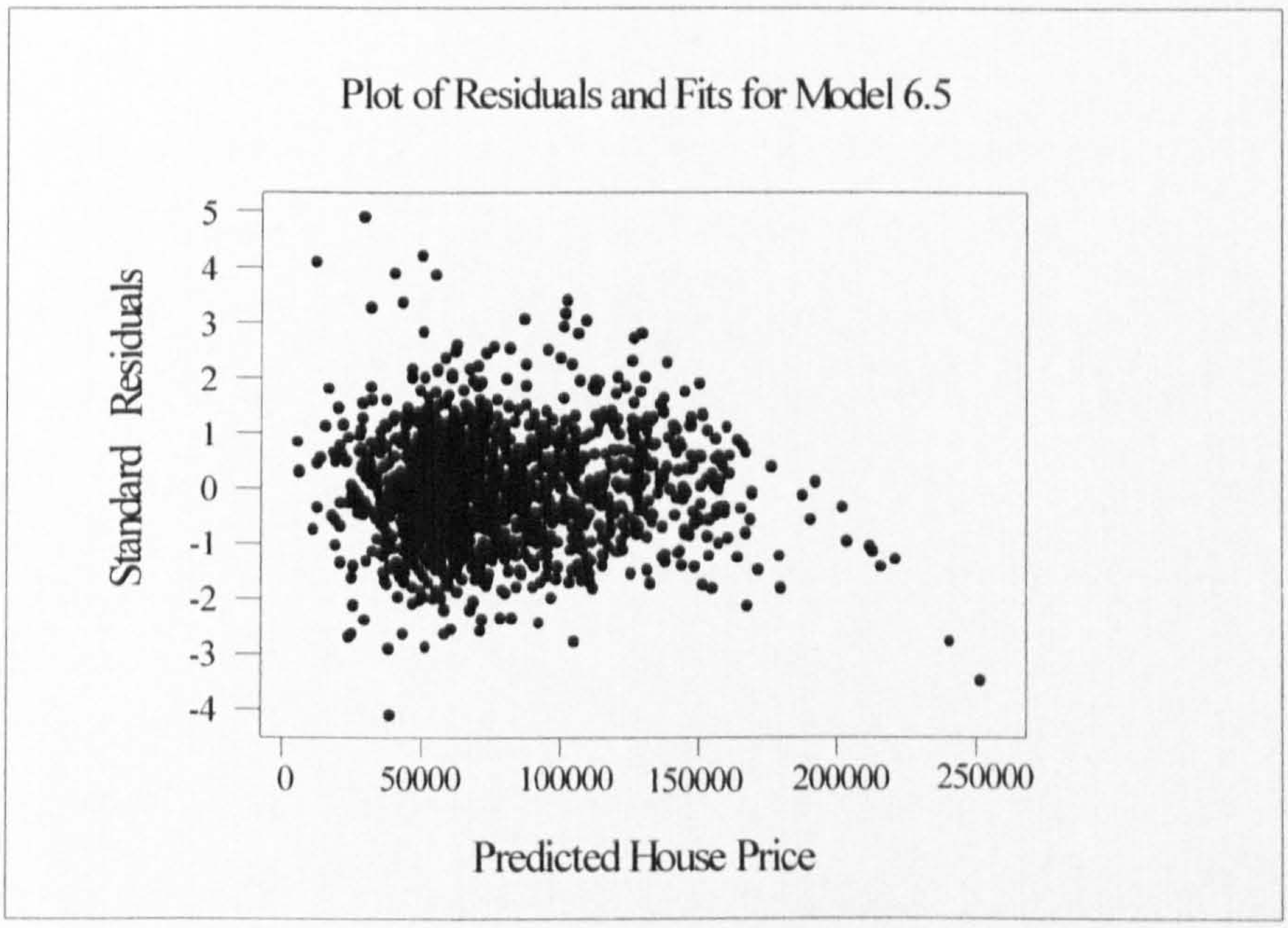
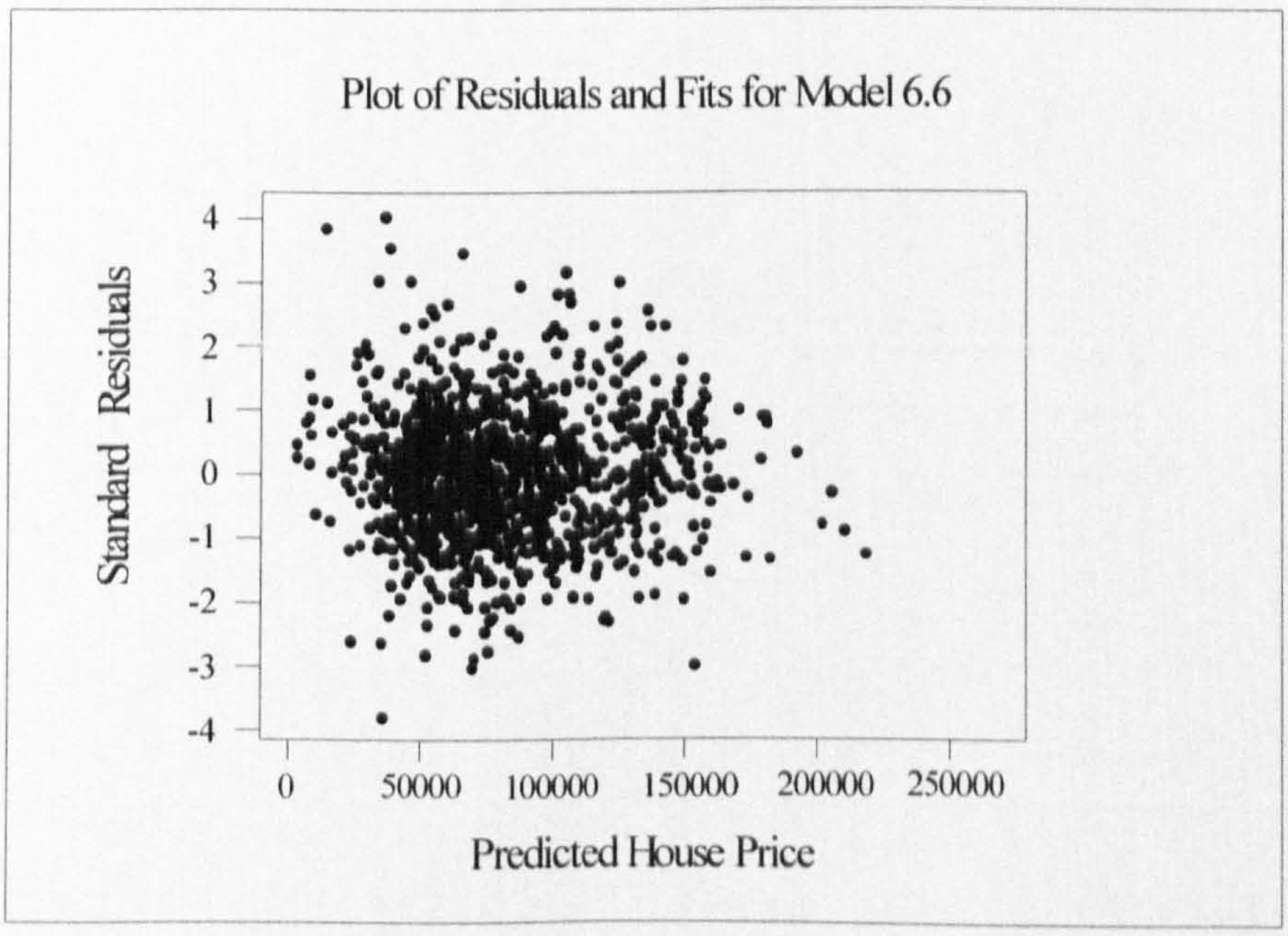


Figure 6.62
Test for Residual Heteroscedastcity





evident, the plots are suggestive of heteroscedasticity, with error variance increasing with price.

A more accurate test for heteroscedasticity that is widely used is the Breusch-Pagan test. This tests the hypotheses that the error variance is a linear combination of the variables. The general strength of the test is that it does not require prior knowledge of the functional form involved, and that there is a computationally convenient means of calculating the test statistics (Kennedy, 1985). The test statistic is based upon the Langrangian Multiplier principle, and is distributed as χ^2 with p degrees of freedom, where p is the number of parameters excluding the constant term. If the test statistic is significant, the presence of heteroscedasticity is assumed. The test is designed to test for the null hypothesis, H_0: $\hat{\sigma}^2 = \sigma^2$, i.e. homoscedasticity. The alternative hypothesis has the following general form of heteroscedasticity: H_1: $\hat{\sigma}^2 = \sigma^2 f(Z)$, where Z is a matrix which incorporates observations on variables that determine the form of heteroscedasticity.

The Breusch-Pagan test was applied to each of the variables in the two Models. Table 6.17 below is a summary of the results for a selection of variables which had particularly large values. These demonstrate that several of the variables in both models are a cause of heteroscedasticity, particularly those variables concerned with house size, such as floor area and number of bedrooms.

Table 6.17.

A Selection of Breusch-Pagan Test Results for Model 6.5 and 6.6

Model	**Variable**	**Test Statistic**
	Floor Area	15.9
Model 6.5	Detached	0.7
	Ave Bedroom Floor Area	13.5
	Ave Recreation Floor Area	8.93
Model 6.6	Ave Kitchen Floor Area	1.03
	Number of Bedrooms	7.80

χ^2 at 95% with 1 degree of freedom = 3.84
χ^2 at 99% with 1 degree of freedom = 6.63

Hence, these tests corroborate Figures 6.62, and it can be concluded that the assumption of identically distributed errors has been violated. Such a result was expected given the problems with the traditional specification of the hedonic model outlined above.

II. Spatial Autocorrelation

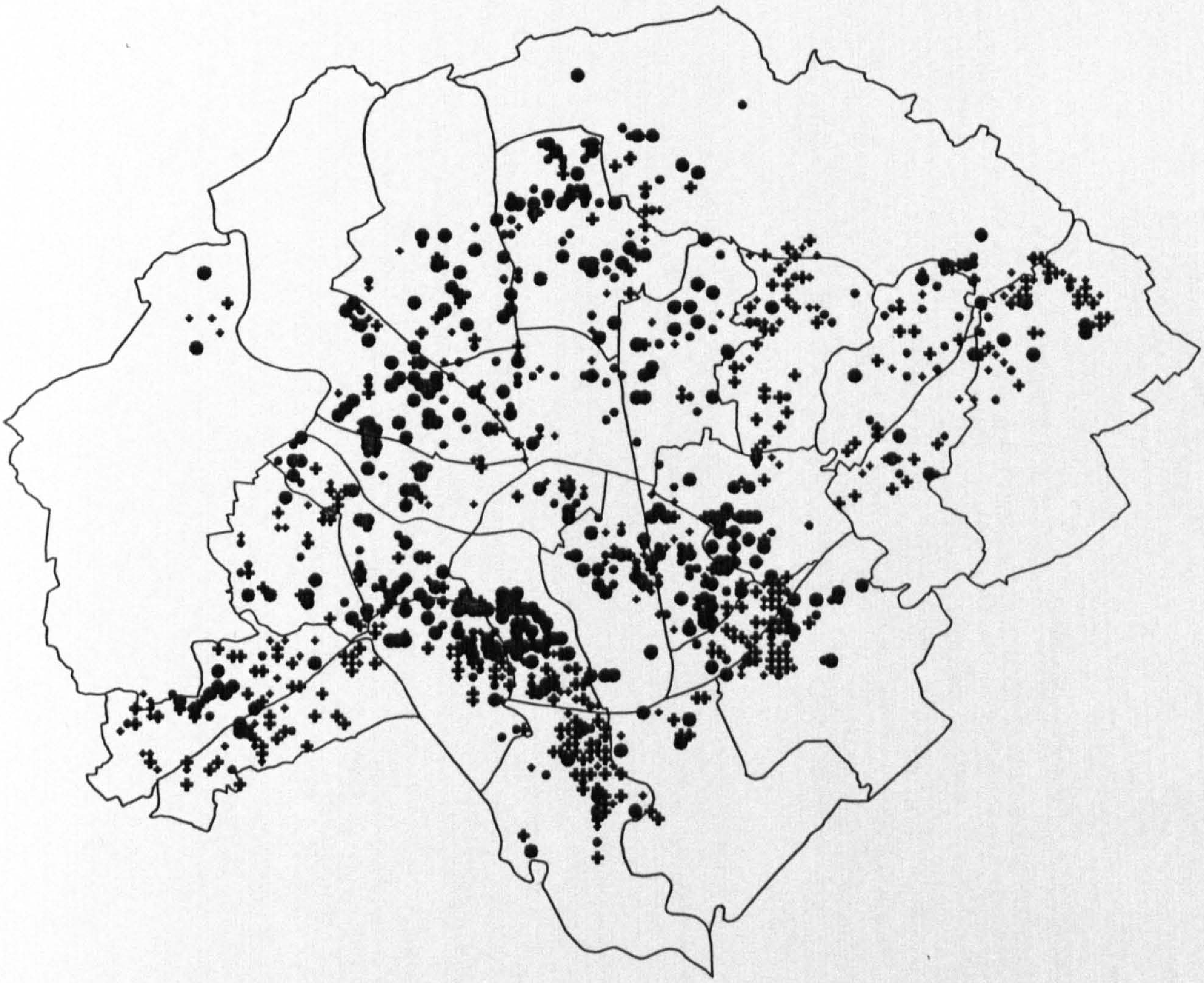
Autocorrelation occurs if the errors are not independent. This can lead to bias in the estimation of residual variance and therefore in measures of the success of the model, such as t-tests, and inefficiency in the estimation of the regression coefficients. In particular, the R-squared value is inflated, suggesting that the model has a better fit than it does. It can appear in the residuals for several reasons: because a linear model has been used for a non-linear relationship; because one or more of the independent variables have been omitted; or because the values of the dependent variable somehow influence each other. The latter case can be conceptualised in terms of dependency across time and space. Dependency across time, when the values of one time period are influenced by the values of the preceding time period, is the best understood form of dependency in the error term, and is referred to simply as autocorrelation. Spatial autocorrelation is dependency across space, and it is this form of dependency that is a cause for concern in the hedonic models.

There are two main causes of spatial autocorrelation in house prices. Neighbouring house prices can influence each other directly via the property valuation process (see *Chapter Two*), or the housing attributes that determine house prices are themselves spatially autocorrelated, such as environmental quality. Cliff and Ord (1981; pp. 141) refer to these two instances as interactive and reactive causes respectively. If all the process which generate house price are solely reactive, and all the causative factors can be found and modelled, then the errors will be completely lacking in spatial autocorrelation. Any spatial autocorrelation that does occur can be attributed to omitted reactive variables. If interactive processes are present, then the residuals will be spatially autocorrelated. The processes that cause spatial autocorrelation may operate in both directions and in two or more dimensions.

If the neighbouring errors are similar across space, then they are said to show positive spatial autocorrelation. Conversely, negative spatial autocorrelation exists when errors close together in space tend to be more dissimilar than errors which are further apart. Zero spatial autocorrelation occurs when the errors are independent of location. Furthermore, it was discussed in *Chapter Three* that scale is implicit in the definition of spatial objects, particularly if these are arbitrarily defined. The corollary of this is that spatial patterns may possess quite different forms of spatial autocorrelation at various scales. Thus the degree of spatial autocorrelation is very much dependent upon scale, and any measure of spatial autocorrelation must be specific to a particular scale. This suggests that it would be

Figure 6.63

The Spatial Distribution of the Standard Residuals of Model 6.5



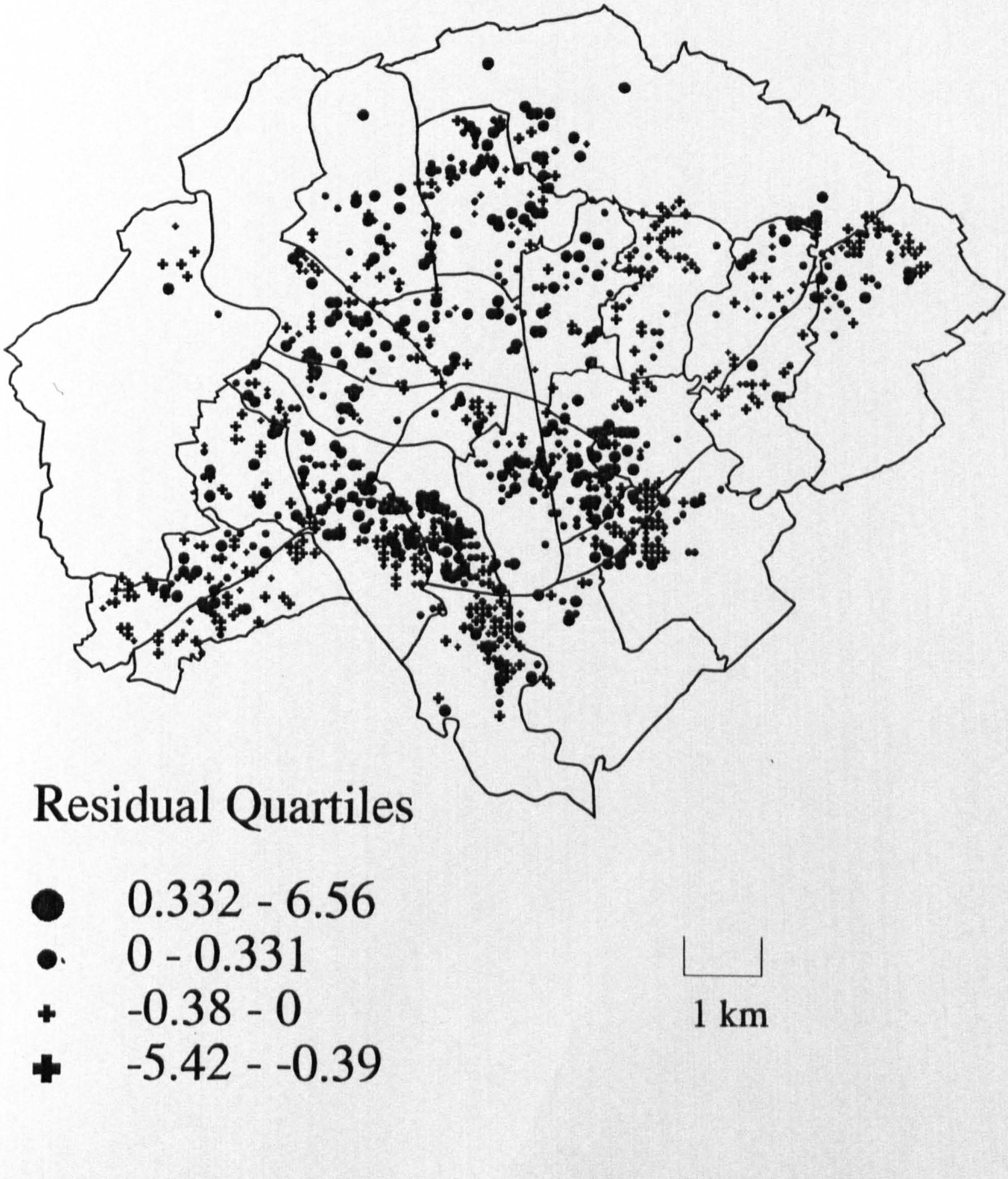
Residual Quartiles

- 0.598 - 4.933
- 0 - 0.597
- -0.636 - 0
- -4.109 - -0.637

1 km

Figure 6.64

The Spatial Distribution of the Standard Residuals of Model 6.6



appropriate to expect spatial autocorrelation in hedonic model residuals to operate at the property, street and neighbourhood levels.

An elementary method of checking for spatial autocorrelation is by mapping the residuals and checking for spatial patterning. Figures 6.63 and 6.64 are residual maps for Model 6.5 and Model 6.6 respectively, plotted at the level of the individual property. Both maps are very similar, and show definite spatial patterning, indicative of positive spatial autocorrelation, which would be expected given the factors that underpin house price determination. Negative residuals are concentrated to the south and east of the city, and also in the peripheral council estates to the west. These are areas were the model has over-predicted house prices. Positive residuals are concentrated in the central Inner Area neighbourhoods and the northern suburbs. Here, the model under-estimates the price of property. This spatial patterning suggests that the hedonic models over-predict house prices in areas of smaller housing, lower social economic class and poorer environmental quality and vice versa. This may imply that locational attributes in the model do not adequately capture locational externalities, although, in conjunction with the tests for heteroscedasticity, it could also imply that the model does not adequately account for property size.

More formal statistical tests for spatial autocorrelation depend upon measures of similarity among attributes and similarity of location. The ways in which the former can be measured depend on the type of data present, while the calculation of spatial proximity depends on the type of object. Finally, there are a number of ways of comparing the two sets of information in compiling a final index, of which the most well known is the Moran Index. The attribute similarity measures used by the Moran Index is analogous to a covariance between the value of the variable at one place and its value at another, whilst a variety of ways have been derived for measuring spatial proximity in order to generate a weight matrix. The Moran Index is positive when nearby errors tend to be similar in attributes, negative when they tend to be more dissimilar, and approximate zero when values are arranged randomly and independently in space. This can be interpreted as a descriptive index, measuring the way things are distributed in space, but at the same time can be seen as a causal process, measuring the degree of influence exerted by something over its neighbours.

A custom written FORTRAN programme was used to calculate a Moran Index (I) for Models 6.5 and 6.6. Since a close inspection of Figures 6.63 and 6.64 shows evidence of neighbourhood clustering of residuals of similar size, a weights matrix was generate based upon neighbourhood location. In this weights matrix, residuals were hypothesized to be spatially autocorrelated only if the properties were located in the same neighbourhood. A two dimensional matrix was created, with a cell containing the value one if two properties were in the same neighbourhood, and zero otherwise. This gave a Moran Index of $I = 0.170$ for Model 6.5 and $I = 0.187$ for Model 6.6. Although both are statistically significant ($I = 0.074$ at 95% with 1421 observations), indicating that the residuals display spatial autocorrelation, Model 6.5 has the lower value.

6.5.3.8 The Standard Hedonic Model

This section has estimated two hedonic models in an investigatory capacity. The remaining objective is to decide which of the final models (Model 6.5 or Model 6.6) best describes the data. This will be based upon the validity of the model with regards the underlying assumptions, and the effectiveness of the model in explaining the variation in house prices. With regards to the former, it can be concluded that although both perform well with respect to the first three assumptions, both models violate the assumptions of independent, identical errors. This issue will be taken up in the subsequent chapters. Therefore, the decision will depend upon the degree to which the models explain house price variation. Although both models explain a high percentage of house price variation, the standard error of Model 6.5 is significantly lower than Model 6.6. Moreover, the measure of floor area performs much better in Model 6.5. This is an important consideration since theoretically, the size of the property will be the most important factor in determining property prices and hence should be modelled correctly. The problems of the floor area variables in Model 6.6 have previously been discussed. Hence, Model 6.6 was dropped in favour of Model 6.5 as the basic hedonic model for the entire Cardiff housing market. This model will be the foundation for the subsequent models in *Chapters Seven and Eight*

Section 6.6 Conclusions

The primary aim of this Chapter has been to describe how the GIS was utilised to generate locational specific attribute data, and then to explore the relationships of this data and the structural attribute data with respect to the Cardiff housing market and its internal structures. Several preliminary hedonic models were also built as a prelude to *Chapter's Seven and Eight*, and the processes behind this model building procedure were discussed in detail. The consequence of this model building exercise was the formulation of a standard hedonic model (Model 6.5) that will be used as the basis in the subsequent macro- and mico-scale studies. Diagnostical and statistical tests demonstrated that the standard model was correctly specified with respect to functional form, multicollinearity and influential observations, but invalidated the assumptions of independent, identically distributed errors. This has been emphasized in *Chapter Two* as a common problem in hedonic models, and is caused by the mis-specification of the hedonic model with respect to spatial data. This issue of specification will be explored in detail in the next chapter.

Chapter Seven

Modelling the Spatial Dynamics of the Cardiff Housing Market

Section 7.1 Introduction

The aim of this chapter is to model the spatial dynamics of the Cardiff housing market. Implicit in this aim is an attempt to incorporate space into the hedonic house price model, using the three specifications developed in *Chapter Two*: the traditional specification, the spatial expansion specification and the multi-level specification. By modelling housing market dynamics within a spatial framework, two key features of the housing market can be addressed. Firstly, that it rarely operates as a unified whole, but rather as a series of submarkets delimited by the housing stock (bundles of housing attributes) and location (geographical areas), and secondly, that the trade-off between housing and transport costs will result in a concave, negative rent gradient from the city centre outwards. Both of these concepts are fundamental to urban economic theory - the former is synonymous with housing market disequilibrium and the latter is synonymous with the bid-rent process (as discussed in *Chapter One*) - but as previous research has shown, empirical evidence concerning these processes is contradictory.

The analysis will begin by building upon the preliminary models estimated in *Chapter Six*, using Model 6.5 as a starting point. The success to which these models capture the spatial structure of the data, and hence the spatial dynamics of the housing market, can be assessed by using diagnostic tests. Spatial effects cause ordinary least squares regression models to exhibit heteroscedasticity and spatial autocorrellation, violating the assumptions of identically distributed and spatially independent errors (see *Chapter Two*). These violations can be identified using the diagnostic and statistical tests described in *Chapter Six*. The variables used in the analysis are summarised in Table 7.1, and have been described in detail in the previous chapter. Briefly, the variables measuring structural attributes have been taken from estate agents sources, whilst the variables measuring locational attributes have been generated from census data, and are acting as a proxy for locational externalities in accordance with previous hedonic research.

Table 7.1
The Cardiff Housing Market Variables

Variable	Attribute Type	Abbreviation
Total Floor Area (sq-ft)	Structural	Floor Area
Semi-Detached	Structural	SD
Detached	Structural	D
Bungalow	Structural	B
Number of Bathrooms	Structural	Baths
Number of Shower rooms	Structural	Showers
Full Central Heating	Structural	Full CH
Number of Garages	Structural	Garages
Off-Road Parking	Structural	ORP
Garden: None	Structural	Gdn:None
Garden: 5 - 50 metres	Structural	Gdn:5-50m
Garden: More than 50 metres	Structural	Gdn:>50m
In Need of Modernisation	Structural	Needs Mods
Distance to CBD	Locational	Dist CBD
Socio-economic Class	Locational	Social
Housing Quality	Locational	H.Qual
ED with Local Authority Tenure > 50%	Locational	LA > 50%

The hedonic models are summarised here in a standard format, an example of which is presented in Table 7.2. The first column contains the variable coefficients in bold, which represent the attribute prices. Below these are the standard errors of the coefficients. The bold values in the second column represent the standardized coefficients, and are calculated by standardizing each variable, and using these in the regression. A standard coefficient of 0.5 implies that a one standard deviation change in the variable will result in a 0.5 standard deviation change in house price, allowing the relative importance of each attribute in the model to be assessed. Below the standard coefficients are the t-statistics. The third column contains the variance inflation factors (VIFs) as a measurement of multicollinearity. A value of greater than 10 indicates that multicollinearity may be a problem in the model. The final column contains the Breusch-Pagan test statistics calculated, as a means of checking for heteroscedasticity. The standard error of the residuals and the coefficient of determination, adjusted for the number of variables in the model, are displayed below each model. For the sake of clarity, the variables that were not significant at the five percent level have been removed from the table.

The remainder of this chapter is divided into five sections. The second section evaluates the success of the traditional hedonic specification in modelling housing dynamics. Section three evaluates models estimated using the spatial expansion method, whilst section four describes the multi-level specification. Section five is concerned with the estimation of the Cardiff rent gradient, whilst the final section concludes with a discussion of the results and their implications for *Chapter Eight*

Section 7.2 The Traditional Hedonic Specification

7.2.1 Introduction

The different hedonic specifications were discussed in detail in *Chapter Two*. Here, the traditional hedonic specification was described as the most common specification, and was used to estimated the preliminary hedonic models in *Chapter Six*. It is the basic specification from which the subsequent specifications are derived. It treats the housing market as a unified whole, with supply and demand schedules held constant across urban space. Thus, there is no spatial variation in implicit prices. Also, the specification regards locational and structural attributes as operating independently, and at the same spatial level. Therefore:

$$P_i = \alpha X_i + \Sigma \beta_k S_{ki} + \Sigma\gamma_q L_{qi} + \varepsilon_i X_i \qquad \textbf{7.1}$$

Where:

$i = 1, ..., N$ is the subscript denoting each property;

P_i is the price of property i;

$k = 1, ..., K$ is the number of structural attributes;

$q = 1, ..., Q$ is the number of locational attributes;

α, β, γ and ε are the corresponding parameters;

X_i is a column vector which consists entirely of ones.

7.2.2 The Initial Model

Before the results of the initial model are analysed, it would be useful to make some *a priori* expectations against which the estimated parameters can be gauged. Firstly, it can be expected that floor area, the dwelling type dummy variables, the number of bathrooms, shower rooms, and garages and the presence of full central heating and off road parking will be positively valued. With respect to the garden size variables, the no garden category is expected to be negatively valued against the base category of a garden up to five metres, whilst the remaining garden variables are expected to be positively valued. The Needs Mods variable is also expected to be negatively valued. In terms of the locational attributes, *a priori* expectations are slightly more ambiguous. In accordance to micro-economic theory, distance to the CBD is expected to be negatively valued, with house prices decreasing with distance. Likewise, the dummy variable measuring whether the ED within which a house is located has over fifty percent Local Authority tenure is also expected to be negative, reflecting the stigma effect associated with Local Authority built housing stock. However, due to their nature, the valuation of the two principal components are slightly more vague. Nevertheless, it is expected that the socio-economic class variable will be positively valued, since positive values represent areas of above average social class, whilst the Housing Quality variable is expected to be negatively valued, since positive scores represents housing stock of below average state of repair overall.

Table 7.2 is a summary of the parameters of Model 7.1, which is identical to Model 6.5. The first thing to note is that all of the attributes have the theoretically correct signs, and the majority are significant at the 1 % level. The R-squared (adjusted) statistic implies that the model is successful in explaining four-fifths of house price variation in Cardiff. Since the data have been deviated around their means, the constant term represents the average price of a typical property in Cardiff (a mid-terraced house with floor size of around 750 square feet), which is estimated as £44,882 [1]. The implicit prices of the attributes reveal some interesting results. As would be expected, the standardised coefficients suggest that the most influential attribute in determining house price is floor area, which has a marginal price of around £35.10 per square foot. Next is the socio-economic class variable, suggesting that

[1] This is a slight departure from the 'conventional' hedonic models of the early 1970s. Here, the constant term was deemed to represent the value of 'land' when all the attributes were set to zero. However, this has always been a dubious assumption (eg. Ball, 1973), since land values and house prices operate within different markets, under different sets of conditions. Furthermore, land values will also be influenced by locational externalities, so it is fallacious to set all the attributes to zero. The notion that house prices could be used to impute land values is an example of how the hedonic house price function was initially used as a means of generating empirical evidence to underpin urban economic theory at that time. The concept that the constant term corresponds to land values has subsequently fallen out of favour

Table 7.2

The Traditional Hedonic Specification
Model 7.1

Predictor	Coeff S.Error	St.Coeff T-stat	VIF	Breusch Pagan
Constant	44882 69.13	0.0557 649.21	*	*
Floor Area	35.10 1.11	0.544 31.73	2.0	15.9
SD	3184 782	0.061 4.07	2.0	0.004
D	16448 1257	0.219 13.08	2.3	0.792
B	15977 1607	0.125 9.94	1.3	0.055
Baths	6478 1186	0.089 5.46	1.4	1.90
Showers	4949 881	0.077 5.62	1.4	42.23
Full CH	4568 773	0.075 6.32	1.2	0.32
Garage	3146 566	0.081 5.56	1.7	0.07
ORP	2825 596	0.065 4.74	1.6	0.26
Gdn:None	-2926 758	-0.051 -3.86	6.1	37.6
Gdn:5-50m	2931 755	0.067 3.88	5.7	13.1
Gdn:>50m	5519 1237	0.071 4.46	4.2	3.08
Needs Mods	-4628 11123	-0.053 -4.16	1.1	0.307
Dist CBD	-1.80 0.170	-0.179 -10.57	2.5	8.64
Social	4077.5 210	0.340 19.42	2.6	60.01
H.Qual	-1389 255	-0.0714 -5.44	1.4	0.449
LA > 50%	3122 1323	0.035 2.36	1.9	65.43

s 13965 R-sq(adj) 83.4%

locational attributes are fundamental in determining property prices. The dwelling type attributes reflect the differences in price that types of housing command after all other housing attributes have been accounted for. The model suggests that there are no significant differences in price *ceteris paribus*, between the base category of end-terrace housing, and linked properties, maisonettes, and flats, since these were insignificant at the five percent level and hence were omitted (see Model 6.3, Table 6.13, *Chapter Six*). However, there are significant price differentials between terraced housing and the other dwelling types, and these can be categorised into two distinct price groups. The first group contains semi-detached housing, which has a modest premium, and does not have great influence in the model. The second group contains detached houses and bungalows. Unlike the former, these construction types both have a notable influence on property prices, increasing the price of a typical property by more than a third *ceteris paribus*. It is also interesting to note the relative importance of the detached dwelling variable. The standard coefficient suggests that its influence in the model is greater than any of the other attributes, with the exception of floor area and socio-economic class. Another unexpected result is the value of a separate shower room (£4,949), which is greater than central heating (£4,568) and a garage (£3,146). Finally, it is worth noting that the value of a property that is in need of modernisation is, on average, £4 628 less than one that is structurally sound. This compares to an estimated mean total repair cost of around £3,200 in the CHCS (see Figure 6.20)

The locational attributes appear to have more intuitive coefficients. Distance from the CBD is significant with the anticipated sign and is a fairly influential attribute of property price. Its implicit price suggests a rent gradient for Cardiff of £1.80 per metre, although this will be examined in more detail in a later section. The standard coefficient of the socio-economic class variable would imply that locational attributes have a significant impact upon house price, relative to the other housing attributes. Since the variable is a principal component, the implicit price of £4,077 is harder to interpret. However, as was described in *Chapter Six*, positive values represent areas of above average socio-economic class and negative values below average socio-economic class, and these have been correlated with factors relating to locational externalities. In areas of average socio-economic class, the variable becomes zero and the model is representative of the typical property in the typical area (the constant term). In areas of high socio-economic class (principal component score of between 1.58 to 3.7) locational effects add an extra £6442 - £15 097 to the value of a typical property. Conversely, in areas of low socio-economic class (principal component score of between -1.28 to -6.74), locational effects reduce the value of a typical property by between £5 219 to £27 482. Housing quality operates upon house prices in a similar fashion, albeit to a lesser extent. Of course, the actual locational externalities that are responsible cannot be determined due to the surrogate nature of the variables, but shall be examined in more detail in *Chapter Eight.*

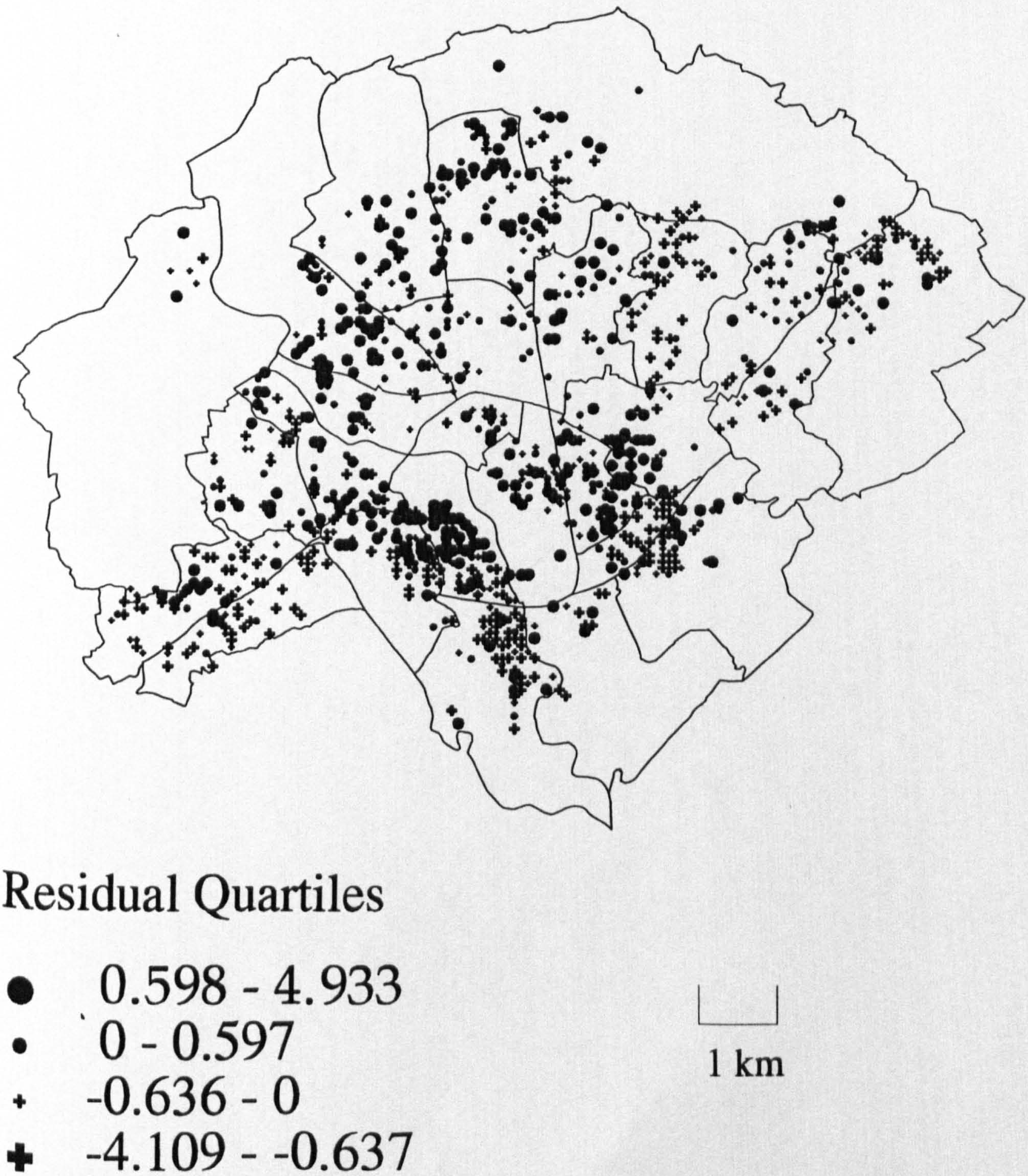
An unexpected result is the implicit price of the Local Authority housing tenure dummy variable. The positive sign would suggest that properties located in areas of predominately Local Authority tenure are marginally more expensive (by £3 122) *ceteris paribus*. This is counter-intuitive since one would expect house prices to be lower due to the stigma element attached to these areas. Therefore, this result could suggest that house prices in these areas are inflated due to restricted availability of houses for sale. Since the variable indicates areas of predominately Local Authority tenure (over 50%), owner occupation would be low and hence the supply of properties constrained. If the demand for housing in such areas is large enough, this restriction in supply would result in an increase in house prices relative to the rest of the city. However, this depends upon a significantly large and localised demand for owner occupied properties in these areas by either in-migration or by new household formation in areas where there are strong existing family ties. The former is doubtful since anecdotal evidence from estate agent sources suggests that the pre-dominant migration patterns are either within the area, which may not create the necessary market conditions, or out of the area. In-migration by new owner occupiers rarely occurs. The alternative explanation is that the model has been mis-specified. This is examined in a later section.

7.2.2.1 Testing for Spatial Effects

Diagnostic tests were performed upon the residuals of each model in order to evaluate their effectiveness in modelling the spatial structures of the housing market. *Chapter Six* demonstrated that heteroscedasticity and spatial autocorrelation were inherent in the initial models. This is now examined in more detail. The Breusch-Pagan test statistics indicate that several variables suffer from heteroscedasticity. With respect to the structural attributes, floor area, the detached housing variable, the number of shower rooms and garden size all display heteroscedasticity. This may be indicative of spatial parameter drift, and omitted variables (Can, 1992), although the preliminary data analysis concluded that floor area may also suffer from non-constant variance. Variables that are not heteroscedastic are those attributes that are restricted to a small subset of housing bundles and locations, such as large gardens.

Figure 7.1

Standard Residuals of Model 7.1



Chapter Six described the procedures for testing for autocorrelation, and concluded that the initial model suffered from spatial autocorrelation, possibly due to a mixture of heteroscedasticity, omitted variables - particularly those pertaining to locational externalities - and the failure of the hedonic model to account for the spatial dynamics of the housing market. This is evident in Figure 7.1, in which the spatial patterning of residuals reveals a non-random distribution, with evidence of community clustering of residuals of similar size. This was confirmed with a significant Moran I test for spatial autocorrelation ($I = 0.170$). *Chapter Six* concluded that the model over-predicts property prices in areas of smaller housing, lower socio-economic class and poor environmental quality and vice versa.

7.2.2.2 Summary

The initial model highlights some interesting results. Firstly, its adjusted R-squared statistic suggests that it explains the variation in house price remarkably well. Secondly, unexpected results, such as the large implicit price of structural attributes for showers, suggest problems with the model specification, and this is confirmed by the indication of heteroscedasticity and spatial autocorrelation in the residuals. Although this is in some way related to the poorly specified locational variables, and the spatial misspecifications associated with the traditional hedonic specification, it may also be indicative of missing structural interaction terms caused by a failure to model submarkets for individual housing bundles. This is examined in the next section.

7.2.3 Modelling Submarkets: Housing Bundles and Structural Interactions

7.2.3.1 Introduction

The traditional specification asserts that the implicit price of a structural attribute is constant across bundles of housing attributes, such that the price for a unit of floor area is identical for terraced housing as for detached housing. However, Rosen (1974) has argued that these housing attribute bundles cannot be untied and repackaged to reflect the consumers desired mixed of attributes. This implies that the available mix of internal structural attributes of a housing bundle is limited and constrained. For instance, it is unusual for a terraced house to have more than four bedrooms, and a detached house to have less than three, although such a combination is theoretically allowed by the traditional specification. Since different groups of households will desire different mixes of attributes, housing bundles may have

Table 7.3

The Traditional Structurally Expanded Specification
Model 7.2

Predictor	Coeff S.Error	St.Coeff T-stat	VIF	Breusch Pagan
Constant	47425	0.051	*	*
	79.36	597.59		
Floor Area	36.60	0.535	2.1	14.2
	1.25	29.26		
Floor SD	2.90	0.053	2.2	1.62
	0.99	2.92		
Floor D	13.0	0.203	3.4	0.76
	1.45	8.96		
Floor B	18.70	0.139	1.6	1.02
	1.95	9.59		
Full CH	5689	0.083	1.2	0.05
	767	7.42		
Garage	4025	0.095	1.7	0.005
	606	6.64		
ORP	2947	0.064	1.6	0.043
	645	4.57		
D Shower 1	6827	0.060	1.7	2.08
	1712	3.99		
FPB Shower 1	15978	0.05	1.1	13.4
	3234	4.94		
D Baths 2	7337	0.041	1.4	0.49
	2556	2.87		
B Baths 2	-10921	0.034	1.2	0.91
	4044	-2.70		
Gdn:None	-4238	0.071	1.5	34.4
	811.9	-5.22		
Gdn:5-50m	3343	0.075	2.5	18.71
	804	4.16		
Gdn:>50m	5280	0.065	2.0	7.76
	1354	3.90		
Needs Mod	-5379	0.055	1.1	0.19
	1537	-3.50		
Dist CBD	-1.90	0.172	2.4	5.4
	0.19	-9.93		
Social	4298	0.337	2.6	48.9
	225	19.09		
H.Qual	-1504	0.072	1.4	0.8
	274	-5.49		
LA > 50%	2789	0.03	1.9	51.0
	1416	1.97		

s 13356 R-sq (adj) 85.40%

different supply and demand schedules operating upon them. This may result in the variation in the implicit price of structural attributes between housing bundles. If this is the case, then using a single variable to measure the implicit price of an attribute may result in heteroscedasticity due to non-constant variance, since the single variable is measuring the effect of several (omitted) variables. In addition, the coefficient of the single variable will represent the weighted average of the implicit prices of the omitted attributes, and not the true implicit price of the attribute. However, with the exception of Schnare & Struyk (1976), there has apparently been very little or no work upon modelling housing bundles, despite the concept being fundamental in the housing market literature (see *Chapter One*).

To account for this, Model 7.1 was expanded to allow the distinct housing bundles within Cardiff to be explicitly modelled. Since it is usual to describe a property by its dwelling type, and each dwelling type embodies a typical set of structural attributes (see *Chapter Six*), this was used to identify each housing bundle. In Model 7.1, dwelling type is regarded as an additional premium on the price of a mid-terrace. But significantly higher implicit prices are commanded for detached housing and bungalows. This would be expected if the dwelling type dummy was capturing the effects of omitted structural variables. A similar argument can be applied to the estimated implicit prices for shower rooms, since this attribute tends only to be available within certain housing bundles. Hence, Model 7.1 was expanded by interacting the internal structural variables with the dwelling type dummy variables. These interaction terms represent the previous omitted variables. In addition, the continuous variables measuring the number of bathrooms and garages were converted into several dummy variables, since despite the findings in *Chapter Six*, there is no reason to expect the implicit price of these attributes to increase at a constant rate. Also, with few exceptions, all the properties in the sample had at least one bathroom, so the bathroom variable was capturing the price differentials of those properties with two or more. Since no explicit spatial variables were included in the expansion equations, the resulting model was termed the traditional structural expansion model.

7.2.3.2 Model 7.2 - The Structurally Expanded Specification

Table 7.3 is a summary of the parameters of this model. Again, the constant term is the price of the typical property. As hypothesized, the implicit prices of the internal structural attributes varied with dwelling type. Generally, the structural attributes for terraced properties, linked properties, maisonettes and flats all had similar implicit prices, and did not significantly vary with respect to each other, and hence have been omitted. This suggests that the structural attribute mix for these housing bundles are very similar. Hence, the implicit price of floor area for all these properties is £36.60 per square foot. The model illustrates how this value varies between the remaining dwelling types. For instance, the model implies that the price of floor area for detached housing and bungalows are a third and a half more expensive respectively than flats, maisonettes, terraces and linked properties ((£36.60 + £13.0) = £49.60 and (£36.60 + £18.70) = £55.30 per square foot.). The variance inflation factors (VIFs) for each variable illustrate that the addition of the floor area interaction terms has not lead to a noticeable increase in multicollinearity in the model.

The standard coefficients suggest that floor area is still the most influential attribute in explaining the variation in house prices, although this has now been distributed between floor area in detached houses and bungalows. A surprising result is the negative implicit price for bungalows with two bathrooms. This is counter-intuitive since it suggests that bungalows with two bathrooms are significantly cheaper (by approximately £10 900) than bungalows with one bathroom. Unless there is a structural explanation (such as two bathroomed bungalows being located in less desirable areas, or are smaller in overall size), these results imply that the model is mis-specified.

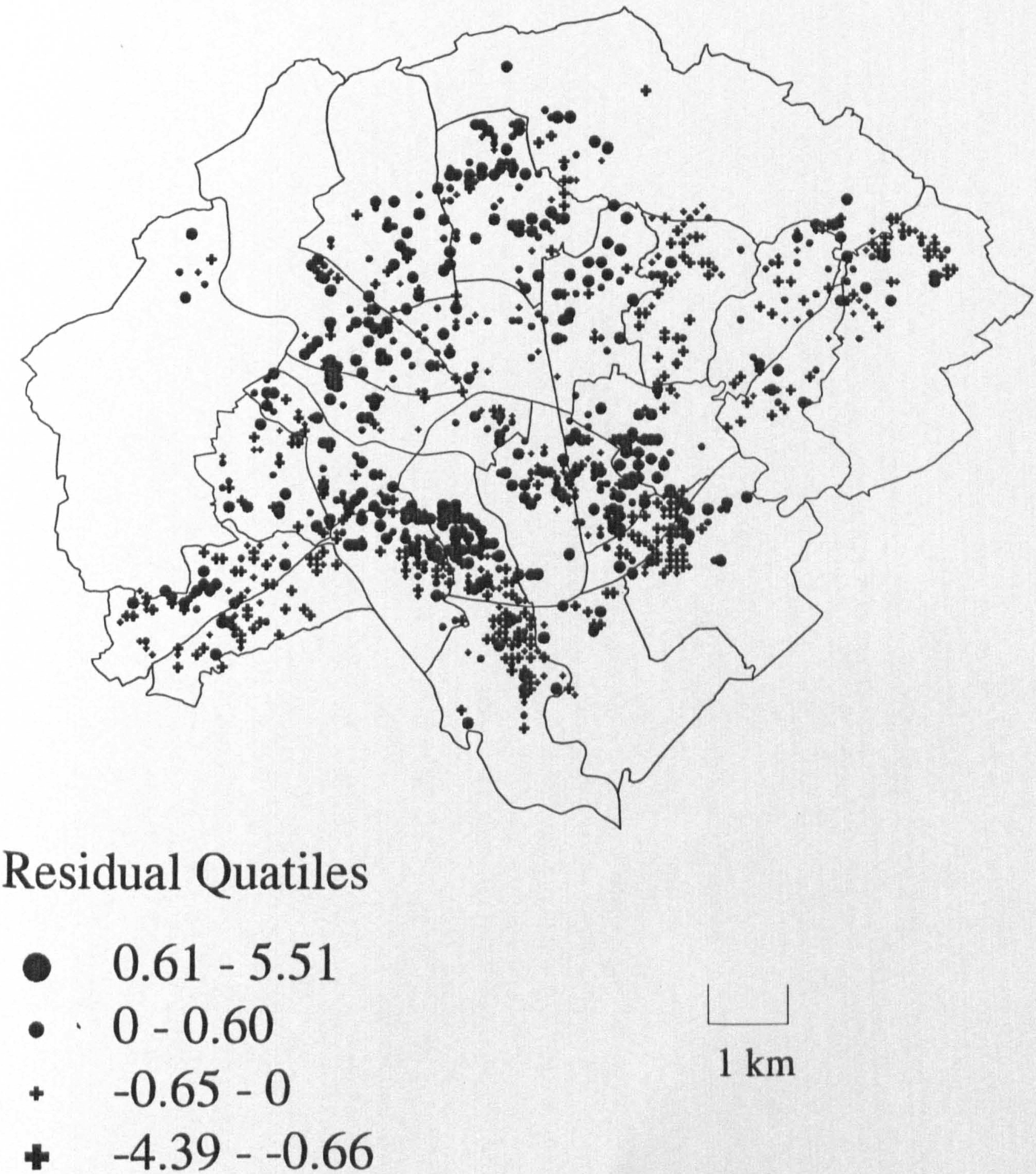
The locational attributes are still highly influential in the model. The social class of an area is still the second most important attribute influencing house price, whilst the accessibility attribute now suggests a rent gradient of £1.90 per metre. However, the Local Authority tenure dummy variable still predicts an additional premium for housing in areas of predominately Local Authority owned stock. Finally, the overall standard error has been reduced, indicating that the model is a better predictor of house price than in Model 7.1 (the adjusted R-squared has increased to 85.4%).

7.2.3.3 Testing for Spatial Effects

The Breusch-Pagan test statistics reveal that heteroscedasticity is still prevalent in the floor area attribute, and has not decreased significantly with respect to the previous model. However, the shower room attribute has seen a marked reduction in heteroscedasticity, and is now only a problem with respect to purpose built flats. The introduction of the structural attribute interaction terms has increased heteroscedasticity in the garden size attributes, although it has had a marked reduction in the locational attributes . These results suggest

Figure 7.2

Standard Residuals of Model 7.2



that it was correct to re-specify the model, and that differentiation between housing bundles is an important structural feature of the Cardiff housing market. The existence of recorded heteroscedasticty suggests that spatial parameter drift may be present. Figure 7.2 revealed that spatial autocorrelation is still a problem, although the Moran Index indicates a decrease ($I = 0.167$) compared to Model 7.1.

7.2.4 Conclusion

The investigation into the traditional hedonic specification has illustrated the importance of diagnostic tests for testing for the existence of heteroscedasticity and spatial autocorrelation, which has been neglected in the literature. The traditional specification suffered from significant heteroscedasticity, and this could be attributed to two main sources: omitted interaction effects between the structural attributes, and suspected spatial parameter drift. The former can be regarded as the failure of modelling submarkets for housing bundles, which was corrected. The latter is a failure to take into account spatial submarkets and shall be investigated in the next section. Both models also suffered from significant spatial autocorrelation, although again this was reduced by including the additional structural variables in Model 7.2. However, mapping evidence suggests that the autocorrelation is linked to communities, which again may be indicative of unaccounted for submarket processes.

The models also highlighted the importance of locational attributes in determining house price, a contentious issue in much previous work. A rent gradient of around £1.85 per metre was estimated, and this figure appeared to be robust between the models. Perhaps the most significant result is the importance of socio-economic class in explaining house price differentials. After floor area had been taken into consideration, this is the second most important explanatory variable as measured by the standard coefficients. The only counter-intuitive result is the premium attached to properties located in pre-dominantly Local Authority owned areas and also the negative value associated with bungalows with two bathrooms.

Section 7.3 Spatially Expanding the Fixed Parameters: The Spatial Parameter Drift Specification

The traditional specification assumed that the Cardiff housing market operates as a unified whole, and as such, that housing attribute prices are invariant across space. However, this is a dubious assumption, given that housing market disequilbrium may lead to the formation of submarkets and spatially varying implicit prices. To capture the presence of submarkets, the traditional hedonic specification was expanded using spatially referenced variables. As was explained in *Chapter Two*, two expansion methods are considered which generate hedonic specifications that incorporate space. The first expands the fixed parameters of the traditional hedonic specification, and is discussed below. The second expands the random parameters, and is examined in Section 7.4.

7.3.1 Introduction

The spatial parameter drift specification models the spatial dynamics of the housing market by expanding the fixed parameters of the traditional hedonic specification. Specifically:

$$P_i = \Sigma (\alpha_0 + \alpha_1 Z) X_i + \Sigma (\beta_{k0} + \beta_{k1} Z) S_{ki} + \Sigma (\gamma_{q0} + \gamma_{q1} Z) L_{qi} + \varepsilon_i X_i \qquad \textbf{7.2}$$

Where Z is a measure of location.

This specification is similar to the discrete space expansion equation (eq. 2.19) as described in *Chapter Two*, and can be regarded as interacting structural attributes with a measure of location. It is hypothesized that it was the omission of these interaction terms that was the cause of the residual heteroscedasticity in the traditional specification. This specification was operationalised by permitting the implicit prices of the housing attributes to interact with the social class variable (see eq. 2.21, *Chapter Two*). In such a specification, there is no implicit price for social class *per se*. Instead, social class can be conceived as driving the implicit prices of the structural attributes across space. The variable was chosen since it can be argued that the valuation of a bundle of housing attributes will be determined in part by the income of the buyer, and this can be proxied by the social class variable. Moreover, since social class is also a proxy for locational attributes, this specification will illustrate how structural attributes vary with locational context.

It may be hypothesized that implicit prices of structural attributes will be more expensive in areas of relatively high socio-economic class *ceteris paribus,* than in areas of lower socio-economic class. This is supported by the significance of the social class variable in determining house price in the earlier models. Two expansion equations were specified. Firstly, a linear equation was specified, implying that changes in implicit prices were proportional to changes in socio-economic class.

$$\delta_k = \delta_{k0} + \Sigma\, \delta_{k1j}\, Z \qquad 7.3$$

Where Z is now social class.
δ represents the fixed parameters in the model.

Secondly, a non-linear quadratic equation was specified, implying that the greatest shifts in attribute implicit prices occurs in areas of very high or very low socio-economic class.

$$\delta_k = \delta_{k0} + \Sigma\, \delta_{k1j}Z + \Sigma\, \delta_{k2j}Z^2 \qquad 7.4$$

The use of higher power polynomials were also considered. But since these may cause interpretative problems, as well as problems of multicollinearity, they were rejected. For reasons of clarity, insignificant drift variables have been omitted from the subsequent tables.

7.3.2 The Traditional Spatial Drift Specification

7.3.2.1 Model 7.3 The Linear Model

The fixed parameters in Model 7.1 were expanded using the linear expansion equation (equation 7.3), and the model was re-estimated.. The average price of the typical property was estimated as £46 812, and this now varies by the social class of an area (Table 7.4). In areas of above average social class (1.58 - 3.7), this price increases by around £6647, whilst in areas of below average social class (-1.28 - -3.7), it is around £10,000 cheaper. This can be used to estimate the additional price of locating in a specific community. Table 7.5 is a summary of the overall social class for each community in Cardiff. Using these values, the average price of a typical property can be allowed to vary. This suggests that the most expensive communities are Radyr & St. Fagans, Lisvane & St. Mellons and Cyncoed, with

Table 7.4
The Linear Spatial Paramter Drift Specification
Model 7.3

Predictor	Coeff S.Error	St.Coeff T-stat	VIF	Breusch Pagan
Constant	46812	0.06	*	*
	73.84	633.92		
Z.Constant	2518	0.20	1.0	9.1
	468.9	5.37		
Floor Area	35.58	0.529	1.9	21.32
	1.12	31.66		
Z.Floor Area	2.99	0.206	3.2	2.63
	0.61	4.92		
SD	2637	0.04	2.0	0.017
	782	3.37		
Z.SD	740	0.04	2.5	1.3
	338	2.19		
D	18393	0.23	1.7	0.49
	1339.6	13.73		
B	16371	0.12	1.4	0.79
	1627	10.06		
Baths	8365	0.11	1.5	4.3
	1262	6.63		
Z.Baths	-2814	-0.09	1.4	0.013
	575.5	-4.89		
Showers	5858	0.09	1.6	11.86
	934	6.27		
Z.Showers	-2225	-0.10	1.9	0.164
	407.5	-5.46		
Full CH	4501	0.07	1.2	0.01
	721	6.24		
Garage	3234	0.08	1.8	1.23
	581	5.56		
Z.Garage	-618	-0.04	2.4	9.1
	273	-2.26		
ORP	2228	0.06	1.6	6.14
	599	3.72		
Gdn:None	-4144	-0.08	1.6	12.6
	770	-5.38		
Gdn:5-50m	2746	0.06	2.5	6.97
	754	3.64		
Gdn:>50m	5989	0.07	2.0	2.15
	1248	4.80		
Needs Mods	-4366	-0.05	1.2	0.005
	1119	-3.90		
Dist CBD	-1.67	-0.13	2.5	1.89
	0.176	-9.44		
H.Qual	-1524	-0.05	1.5	0.51
	266	-5.73		
La > 50%	-10220	-0.14	9.5	2.69
	2542	-4.02		
Z.La > 50%	-4370	-0.17	8.5	3.3
	812	-5.38		

s 9474 R-sq(adj) 84.5%

the least expensive being Butetown, Adamsdown and Ely. These are quite intuitive results given the raw averages in Table 6.8. The relatively large standard coefficient suggests that the way the typically priced property drifts with social class is an important feature of the model.

Table 7.5

Community Ranked By Social Class

Community	Social Class	Community	Social Class
Butetown	-5.20279	Pentwyn	0.145
Adamsdown	-2.83398	Gabalfa	0.16231
Ely	-2.64362	Canton	0.41508
Riverside	-2.44176	Rumney	0.4593
Grangetown	-1.97295	Whitchurch & Tongwynlais	1.46849
Plasnewydd	-1.90944	Roath	1.51733
Splott	-1.69341	Llanishen	1.91029
Trowbridge	-1.36588	Heath	2.32062
Llanrumney	-1.15323	Llandaff	2.49424
Caerau	-1.14426	Rhiwbina	2.8676
Cathays	-0.75659	Cyncoed	2.99627
Llandaff North	-0.45072	Lisvane & St Mellons	3.30597
Fairwater	-0.02188	Radyr & St Fagans	3.528

Similar to the previous models, floor area remained the most important attribute in the model (standard coefficient of 0.529). In areas of average social class, the price of floor area is estimated as £35.58 per square foot. This then varies by £2.99 per square foot as social class deviates from this average. Hence, in areas of relatively high social class, a household would have to pay on average an additional £8.72 per square foot. Conversely, in areas of relatively low social class, the price would be £7.73 per square foot cheaper. Detached housing and bungalows still command a high premium, although this premium does not vary, but remains stable across urban space. However, the premium for semi-detached houses is dependent upon social class, with the price of a semi-detached house increasing as the social class of an area increases. This instability can be explained by the ubiquitous nature of semi-detached houses compared to detached houses and bungalows. The latter tend to be concentrated in areas of similar social class, and hence their implicit prices are less likely to drift.

The only other structural attribute prices that experience drift are baths, showers and garages. In all three cases, the drift coefficient are negative, implying that the additional value of such attributes are increasingly higher in areas of lower social class. This rather counter-intuitive result may be explained by the fact that attributes such as two bathrooms and double garages are rarer in the housing stock in these areas, and hence with sufficient demand, their restricted supply may increase their value. However, it may be the case that additional structural attributes are valued less in areas of higher social class in favour of locational attributes. This will be discussed in a later section.

An interesting result is the implicit price of properties in areas of predominately Local Authority owned housing stock. The average implicit price is negative, as is expected. This price then varies depending upon the social class of the area. Therefore, in areas of average social class, Local Authority housing stock has a negative effect of around £10 000. This negative effect varies proportionately with social class. This is summarised in Figure 7.3, and shows that, for low social class (-6 to -3), properties in areas of predominately Local Authority tenure are more desirable and hence more expensive, than similar properties in non-Local Authority areas. This may be explained by the fact that a high proportion of Local Authority dwellings in Cardiff were built in the inter-war period when construction standards were high (Short, 1981). In particular, they tend to be of better structural quality and built to a lower density than comparable properties in other areas of low social class, which are synonymous with streets of inner city terraces. This combination of structural quality and locational externalities has resulted in the 'stigma effect' only becoming significant in areas of social class of above -3. In these areas, Local Authority housing stock may be viewed with more disdain by the more relatively affluent purchasers, who will also have a better choice of housing stock of comparable quality and size.

Figure 7.3
Local Authority Housing Stock and Social Class: Model 7.3

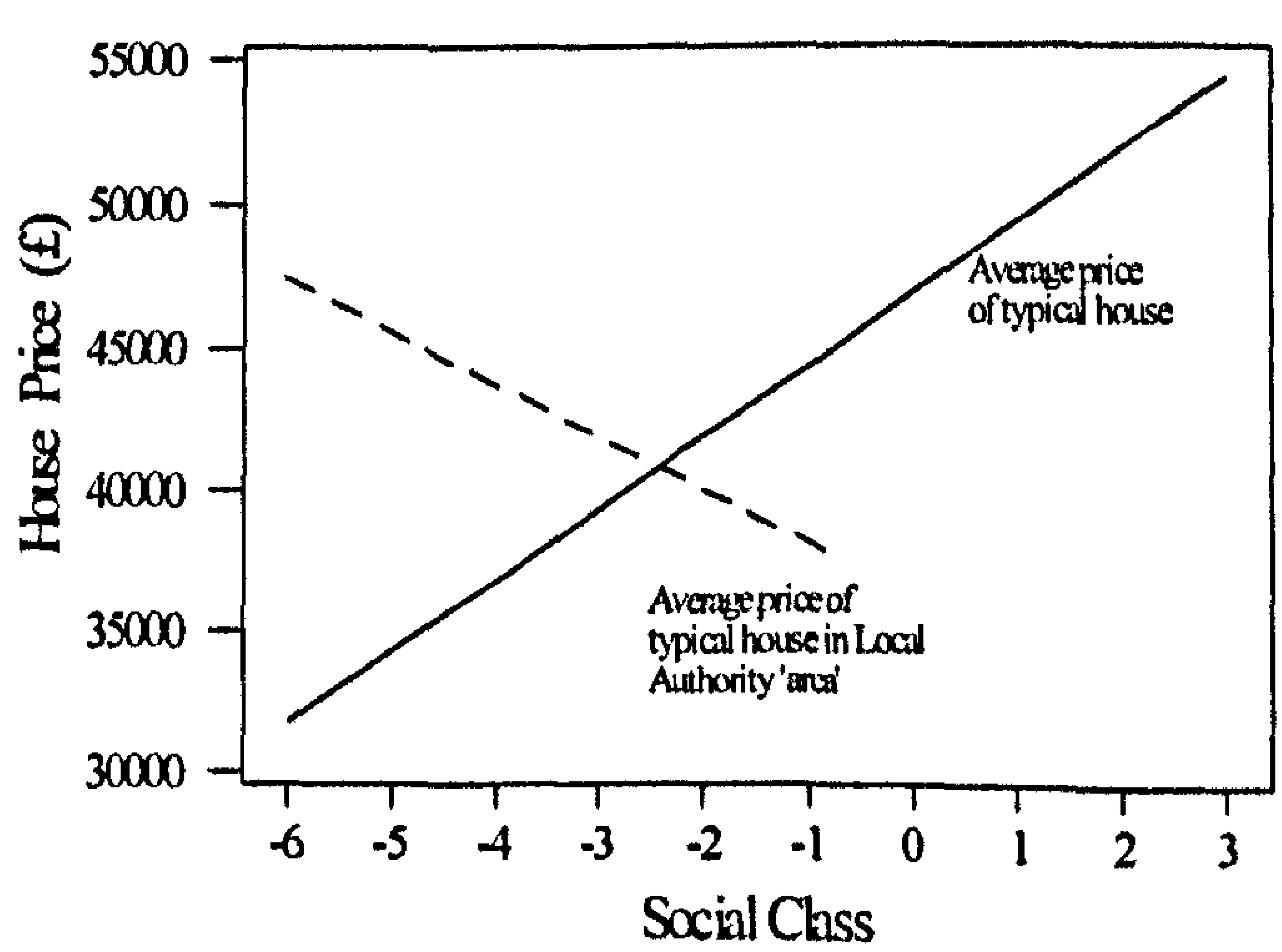


Table 7.6
The Non-Linear Spatial Paramter Drift Specification
Model 7.4

Predictor	Coeff S.Error	St.Coeff T-stat	VIF	Breusch Pagan
Constant	46486	0.06	*	*
	72.46	641.51		
Z.Constant	2614	0.21	1.0	9.7
	470	5.56		
Floor Area	36.77	0.54	2.6	18.53
	1.29	28.60		
Z.Floor Area	2.59	0.202	5.7	3.28
	0.59	4.39		
ZZ.Floor Area	-0.52	-0.04	2.1	0.04
	0.23	-2.23		
SD	2560	0.04	2.0	0.05
	775.6	3.30		
Z.SD	674	0.04	1.7	0.94
	337	2.00		
D	17839	0.22	2.5	0.33
	1308	13.64		
B	16126	0.12	1.4	0.69
	1614	9.99		
Baths	8235	0.11	1.5	3.34
	1253	6.57		
Z.Baths	-2650	-0.04	1.5	0.34
	584	-4.54		
Showers	5885	0.09	1.6	10.9
	927	6.35		
Z.Showers	-2170	-0.10	1.9	0.38
	407	-5.33		
Full CH	4526	0.07	1.2	0.01
	717	6.31		
Garage	2919	0.07	1.7	0.66
	564	5.18		
ORP	2335	0.06	1.6	7.1
	594	3.93		
Gdn:None	-4060	-0.07	1.6	13.14
	765	-5.31		
Gdn:5-50m	2737	0.06	2.5	5.63
	750	3.65		
Gdn:>50m	6032	0.07	2.0	1.75
	1241	4.86		
Needs Mods	-4398	-0.05	1.2	0.001
	1111	-3.96		
Dist CBD	-1.69	-0.13	2.5	1.64
	0.175	-9.61		
H.Qual	-1446	-0.08	1.5	0.16
	262	-5.52		
La > 50%	-9281	-0.13	9.4	3.2
	2543	-3.65		
Z.La > 50%	-3826	-0.15	8.1	3.7
	802	4.77		

s 9408 R-sq(adj) 84.50%

7.3.2.2 Model 7.4 The Non-linear Model

This model presupposes that housing attributes interact with social class in a complex, non-linear fashion. Model 7.3 was re-estimated using the quadratic expansion equation and is summarised in Table 7.6. The principal result of the model is the implication that the spatial variation of floor area is non-linear. The model suggests that not only does floor area vary with social class, but this variation is greater in areas of either very high and very low social class. In both cases, this results in a deduction of £0.52 per square foot. In addition, the garage variables now become stable across the housing market, suggesting that the spatial variation in the linear model was caused by the number of garages acting as a proxy for unaccounted spatial variation in floor area. Finally there is very little difference between the standard errors in the two models, suggesting that the non-linear model does not significantly explain more of the variation in house prices. The remaining variables remain robust.

Figure 7.4 illustrates the way in which the unit price of floor area varies with social class in Models 7.3 and 7.4. It is the non-linear plot which is of interest. This suggests that the unit price of floor area increases at a decreasing rate until it peaks at around £40 per square foot in areas of slightly above average social class (between 1 and 2). The price per square foot then starts to decrease. This result is interesting since it suggests that higher income purchasers value the structural attributes of housing less than purchasers of lower incomes. The corollary of this is that higher income purchasers may be placing greater value upon locational attributes. This is supported by the way the bathrooms and shower room parameters drift, being progressively less expensive in areas of higher socio-economic status. This unexpected feature of the housing market is now examined in more detail.

7.3.2.3 Testing for Spatial Effects

Both models suffer from heteroscedasticity, although this has decreased significantly compared to model 7.1, the non-expanded equivalent. The linear model is more heteroscedastic, with heteroscedasticity concentrated in the floor areas variables, and related attributes such as baths, showers and garden size. This is caused in part by the omitted structural interaction variables, although the garden size variables suggest that perhaps spatial heterogeneity is still problematic. The residual map of model 7.4 (Figure 7.5), and

Figure 7.4

Price per unit Floor Area by Social Class: Model 7.3 / Model 7.4

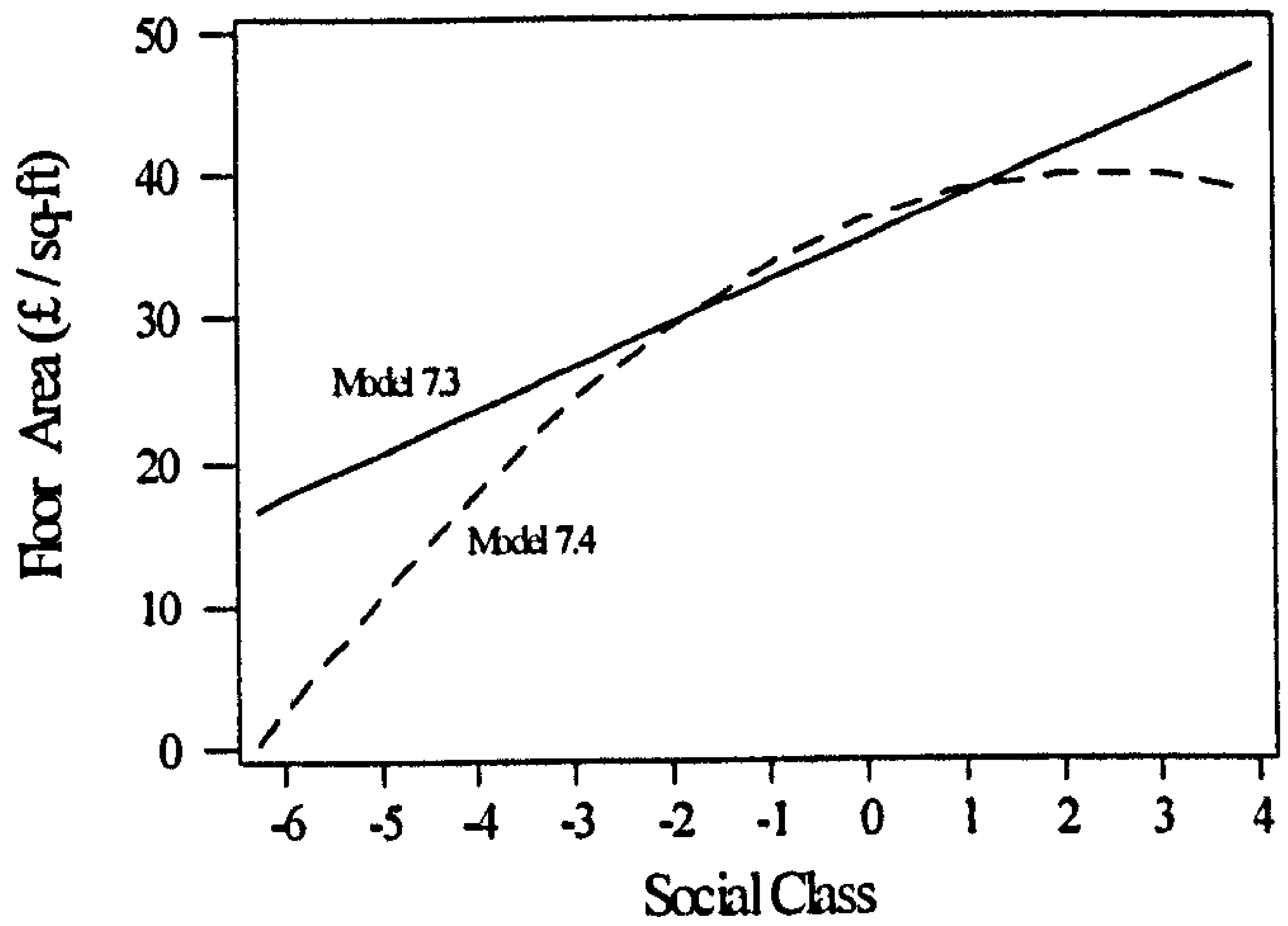


Figure 7.6

Price per unit Floor Area by Social Class: Model 7.5

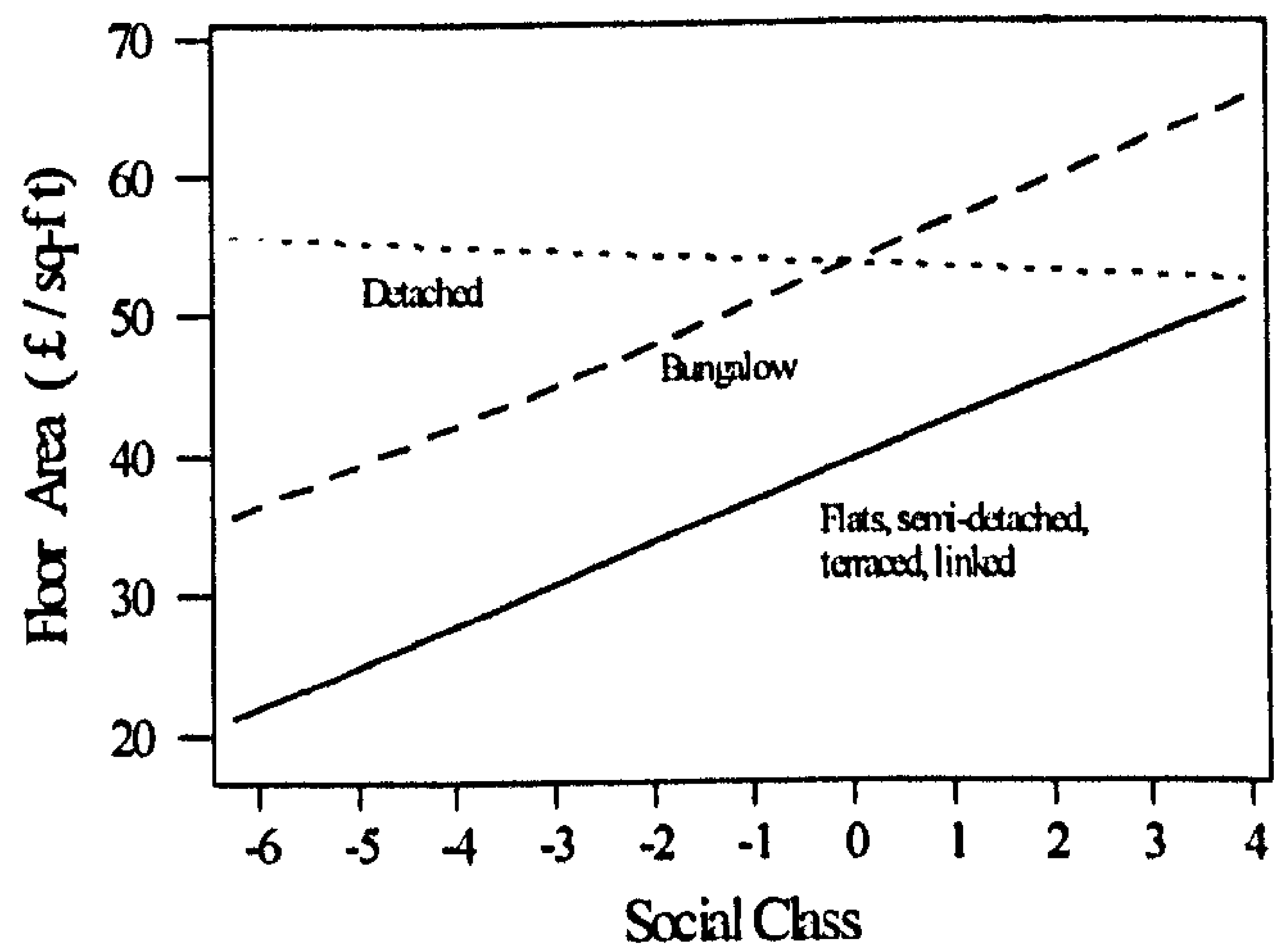
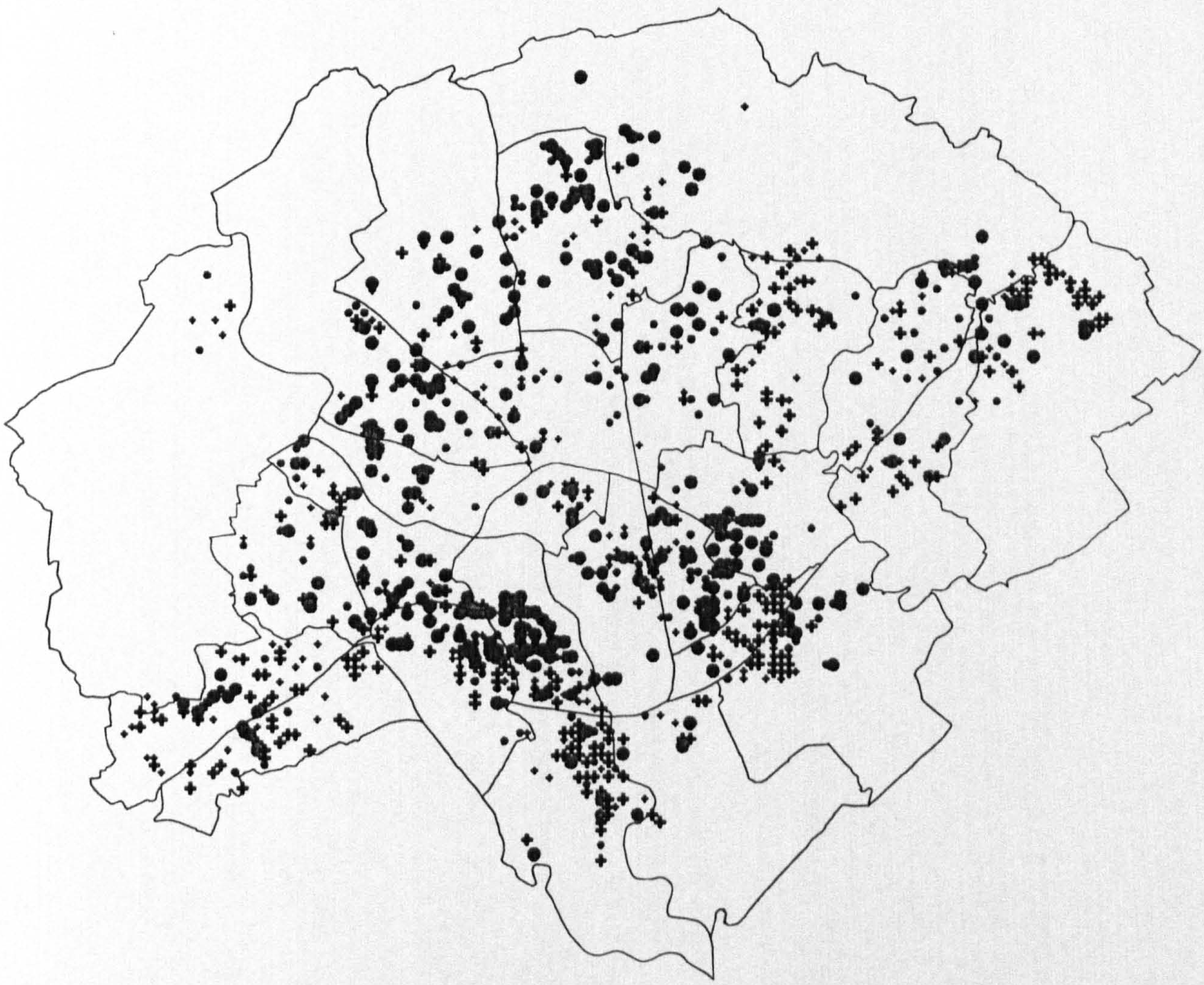


Figure 7.5

Standard Residuals of Model 7.4



Residual Quartiles

- 0.637 - 4.546
- 0 - 0.636
- -0.658 - 0
- -3.833 - -0.659

1 km

Table 7.7
The Structurally Expanded Linear Spatial Drift Parameter Specification
Model 7.5

Predictor	Coeff S.Error	St.Coeff T-stat	VIF	Breusch Pagan
Constant	49518 81.40	0.04 608.30	*	*
Z.Constant	2664 548	0.204 4.86	1.1	10.78
Floor Area	39.33 1.16	0.55 33.83	1.8	15.4
Z.Floor Area	2.91 0.67	0.19 4.31	7.3	2.79
Floor D	14.34 1.61	0.22 8.90	4.1	0.35
Z.Floor D	-3.24 0.69	-0.11 -4.69	3.5	0.00
Floor B	14.42 1.83	0.101 7.88	1.4	0.155
D Bath2	26210 7383	0.12 3.55	8.3	0.08
Z.D Bath2	-5173 2210	0.08 -2.34	8.6	0.055
Z.B Bath2	-5652 2070	-0.04 -2.73	1.2	0.29
D Shower1	7458 1703	0.06 4.38	1.7	1.94
Z.FPB Shower1	-4451 849	-0.06 -5.24	1.4	0.38
Z.SD Shower1	-3569 1399	0.04 -2.55	1.1	2.15
Full CH	5332 782	0.08 6.82	1.2	0.02
Garage	4167 605	0.10 6.89	1.7	0.02
ORP	2979 637	0.06 4.68	1.5	1.5
Gdn:None	-4977 825	-0.08 -6.03	1.6	17.3
Gdn:5-50m	4011 755	0.07 5.31	2.1	9.58
Gdn:>50m	6737 1296	0.08 5.20	1.8	4.78
Needs Mods	-4932 1207	-0.05 -4.09	1.1	0.12
Dist CBD	-1.83 0.19	-0.16 -9.51	2.5	1.9
H.Qual	-1849 284	-0.08 -6.50	1.5	0.82
La > 50%	-11510 2774	-0.13 -4.15	9.8	5.27
Z.La > 50%	-4887 900	-0.18 -5.43	9.1	10.01

s 10230 R-sq(adj) 83.9%

the corresponding significant Moran I test (I = 0.165), both indicate that spatial autocorrelation is still present.

7.3.3 The Structurally Expanded Spatial Drift Specification

7.3.3.1 Model 7.5 The Linear Model

The fixed parameters in Model 7.2 were expanded using the linear expansion equation and the model was re-estimated (Table 7.7). The new model revealed how the internal structural attributes of the different dwellings types drifted with social class. The price of floor area for terraced and linked houses, maisonettes and flats was estimated to be £39.33 per square foot. This then varied by £2.91 per square foot, depending upon social class of the area. Floor area for detached housing has an additional premium of £14.24 per square foot in areas of average social class, with this varying by £3.24 per square foot, although the negative sign implies that the price of floor area per square foot is progressively cheaper as social class increases. Floor area for bungalows also command an additional premium of £14.42 per square foot, although this premium does not drift but rather is constant across Cardiff. This can be explained by the fact that bungalows are concentrated within specific communities (see *Chapter Six*), which have very similar social class. Figure 7.6 summarises this interaction of floor area and social class for each property type. It is interesting to note the negative slope for detached houses and the implication that in areas of high social class, there is very little difference in the price per unit floor area between housing types, with the exception of bungalows. These are the most expensive dwelling type per square foot in areas of above average social class.

The remaining structural attributes reveal an interesting geography of spatial drift. The implicit price of two bathrooms in a detached property is more expensive in areas of lower social class, whilst the implicit price of a two bathroomed bungalow drifts such that they are cheaper in more affluent areas. In the previous, unexpanded model (Model 7.2), this variable had the wrong sign. It would now seem that this was caused by misspecification due to omitted variable bias. In a similar fashion, shower rooms in purpose built flats and semi-detached housing are also cheaper in more affluent areas, whilst those in detached houses are stable across the housing market. The remaining structural attributes are invariant, whilst Local Authority dwelling behave in a similar fashion to the previous model.

Table 7.8
The Structurally Expanded Non-Linear Spatial Paramter Drift Specification
Model 7.6

Predictor	Coeff S.Error	St.Coeff T-stat	VIF	Breusch Pagan
Constant	49222 81.10	0.05 607.03	*	*
Z.Constant	3106 560	0.24 5.55	1.1	12.2
Floor Area	41.34 1.35	0.58 30.54	2.5	12.1
Z.Floor Area	2.42 0.68	0.16 3.56	8.5	5.3
ZZ.Floor Area	-0.55 0.165	-0.06 -3.33	2.4	0.69
Floor D	13.9 1.59	0.21 8.76	4.2	0.25
Z.Floor D	-2.50 0.71	-0.08 -3.53	3.8	1.8
Floor B	14.9 1.82	0.11 8.22	1.4	0.21
D Bath2	22067 7141	0.11 3.09	8.5	0.27
Z.D Bath2	-3807 1667	0.06 -2.27	8.9	0.4
Z.B Bath2	-5409 2048	-0.03 -2.64	1.2	0.17
D Shower1	7487 1682	0.06 4.45	1.7	1.32
Z.FPB Shower1	-4286 840	-0.06 -5.10	1.4	0.15
Z.SD Shower1	-3309 1390	-0.04 -2.38	1.1	2.17
Full CH	5318 773	0.08 6.88	1.2	0.08
Garage	4041 597	0.09 6.77	1.7	0.065
ORP	2934 628.3	0.06 4.67	1.5	1.8
Gdn:None	-4809 816	-0.07 -5.89	1.6	18.7
Gdn:5-50m	3945 746	0.08 5.29	2.1	10.21
Gdn:>50m	6929 1283	0.08 5.40	1.8	5.7
Needs Mods	-4902 1187	-0.05 -4.13	1.1	0.07
Dist CBD	-1.87 0.19	-0.164 -9.80	2.5	1.65
H.Qual	-1764 281	-0.08 -6.27	1.5	0.28
La > 50%	-10448 2786	-0.12 -3.75	9.8	6.9
Z.La > 50%	-4326 907	-0.17 -4.77	9.2	11.8

s 10127 R-sq(adj) 84.0%

7.3.3.2 Model 7.6 The Non-linear Model.

Similar to Model 7.4, floor area was the only variable to drift across space in a non-linear fashion (Table 7.8). This additional parameter has the effect of increasing the spatially varying implicit price of floor area for bungalows, but reducing the implicit price of floor area for all other properties. The remaining parameters appear to be robust with respect to non-linear spatial drift. Figure 7.7 summarises the new patterns of floor area drift. The model now suggests that the implicit of price floor area for all dwelling types increase with social class at a decreasing rate, before peaking and then declining in areas of above social class. The steepest decline is for detached housing, in which floor area is the most expensive in areas of average social class. Again, bungalows are the most expensive properties with respect to size in areas of above average social class.

Figure 7.8 shows the same graph re-plotted using average community social class values. This suggests that, for detached housing, the most expensive housing stock is located in Cathays, Llandaff North and Fairwater, whilst buyers are paying less for housing space in Cyncoed, Lisvane & St. Mellons and Radyr & St Fagans. This may imply that buyers in these areas value locational attributes more, and hence maybe have more incentive to exclude negative externalities from the locality than in areas of more average social class. Finally, Figure 7.9 is a summary of the effect that an area with a high degree of Local Authority tenure has upon property prices. The graph is very similar to Figure 7.3, implying that the results are robust to changes in model specification. Again it would appear that, in areas of low social class, properties are more desirable if they are located in areas of predominately Local Authority tenure, with this desirability declining with increasing social class.

7.3.3.3 Testing for Spatial Effects

Heteroscedasticity was reduced in the floor area variable in both models compared to Models 7.3 and 7.4. However, there was an increase in heteroscedasticity in the garden and Local Authority variables. This heteroscedasticity was slightly greater in the non-linear model. However, these did display a marked reduction in heteroscedasticty compared to Model 7.2, suggesting that spatial heterogeneity was a problem in the traditional specifications. Finally, Figure 7.10 reveals that positive spatial autocorrelation is still

Figure 7.7

Price per unit Floor Area by Social Class: Model 7.6

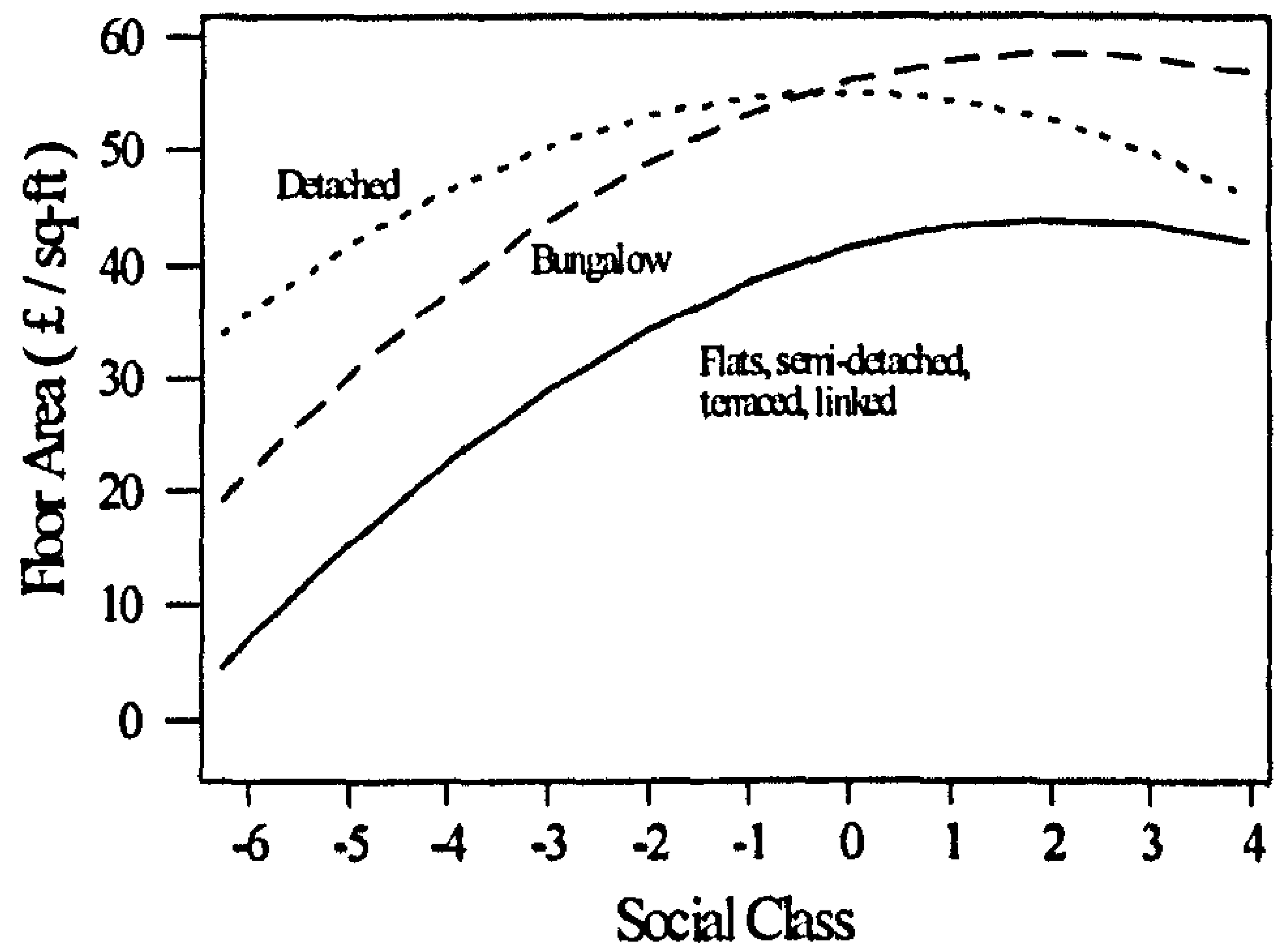
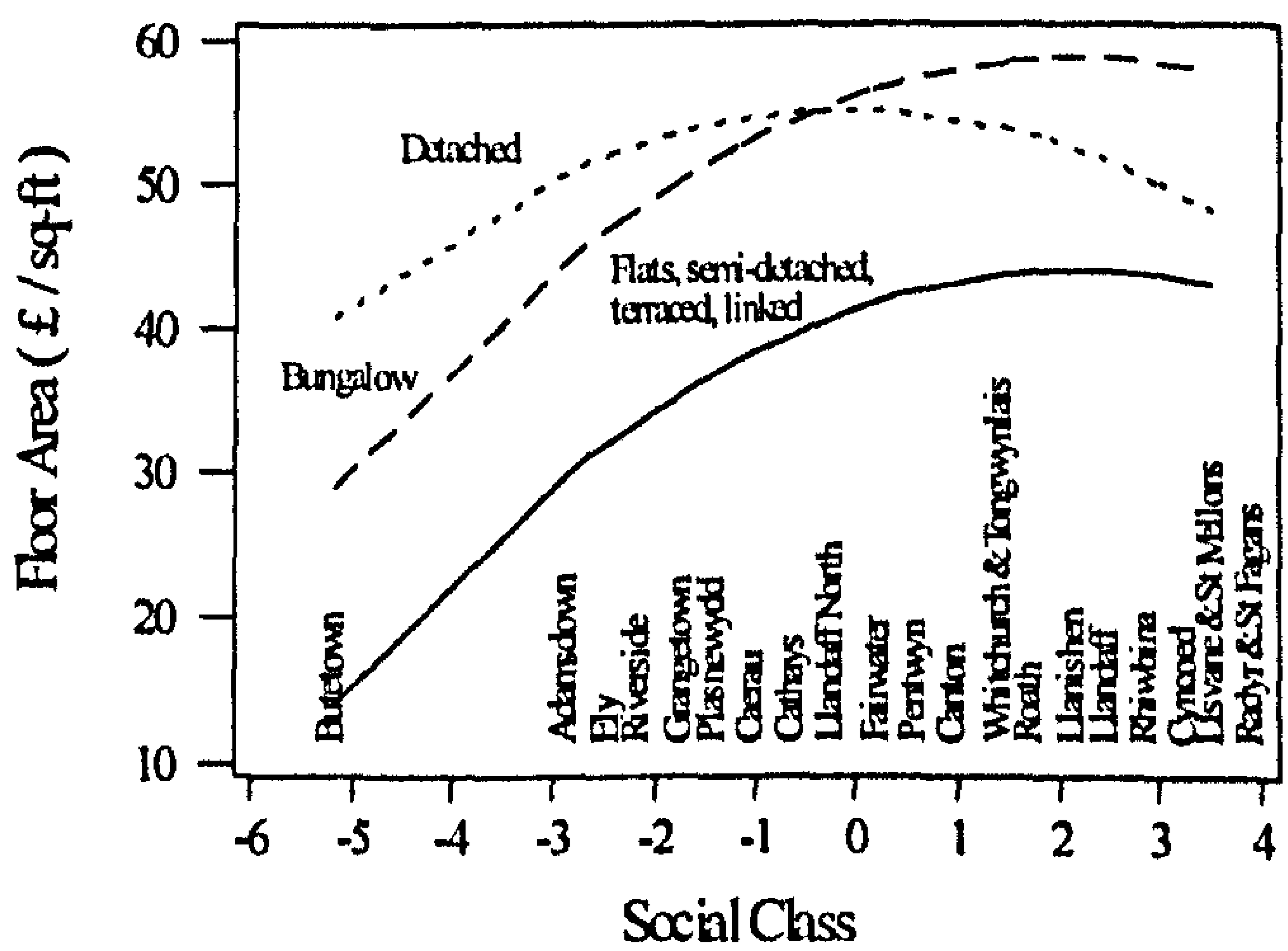


Figure 7.8

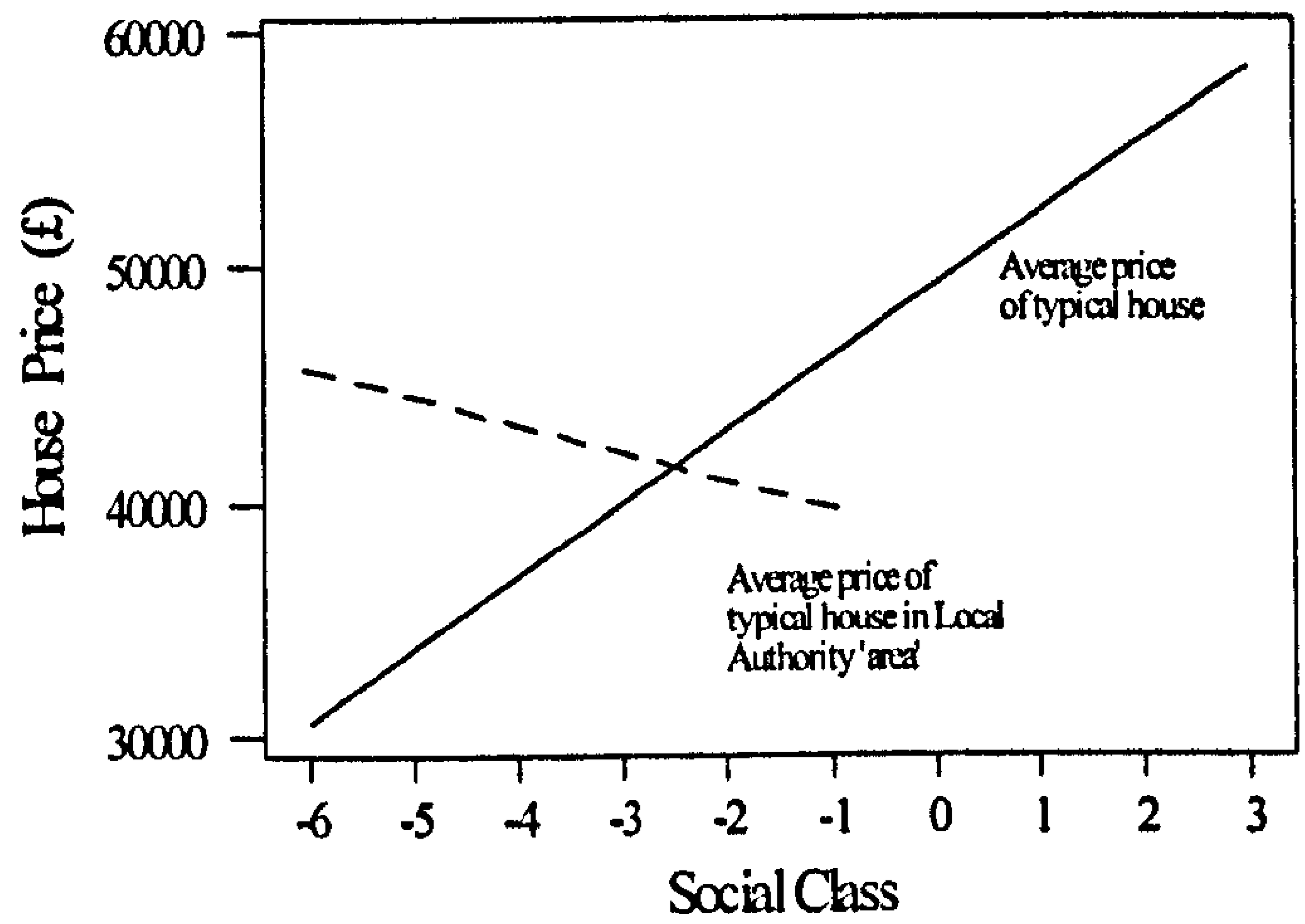
Price per unit Floor Area by Community Level Social Class: Model 7.6



prevalent in Model 7.6, with a significant Moran I Test (I = 0.157), despite the attempt to capture submarket processes.

Figure 7.9

Local Authority Housing Stock and Social Class: Model 7.6

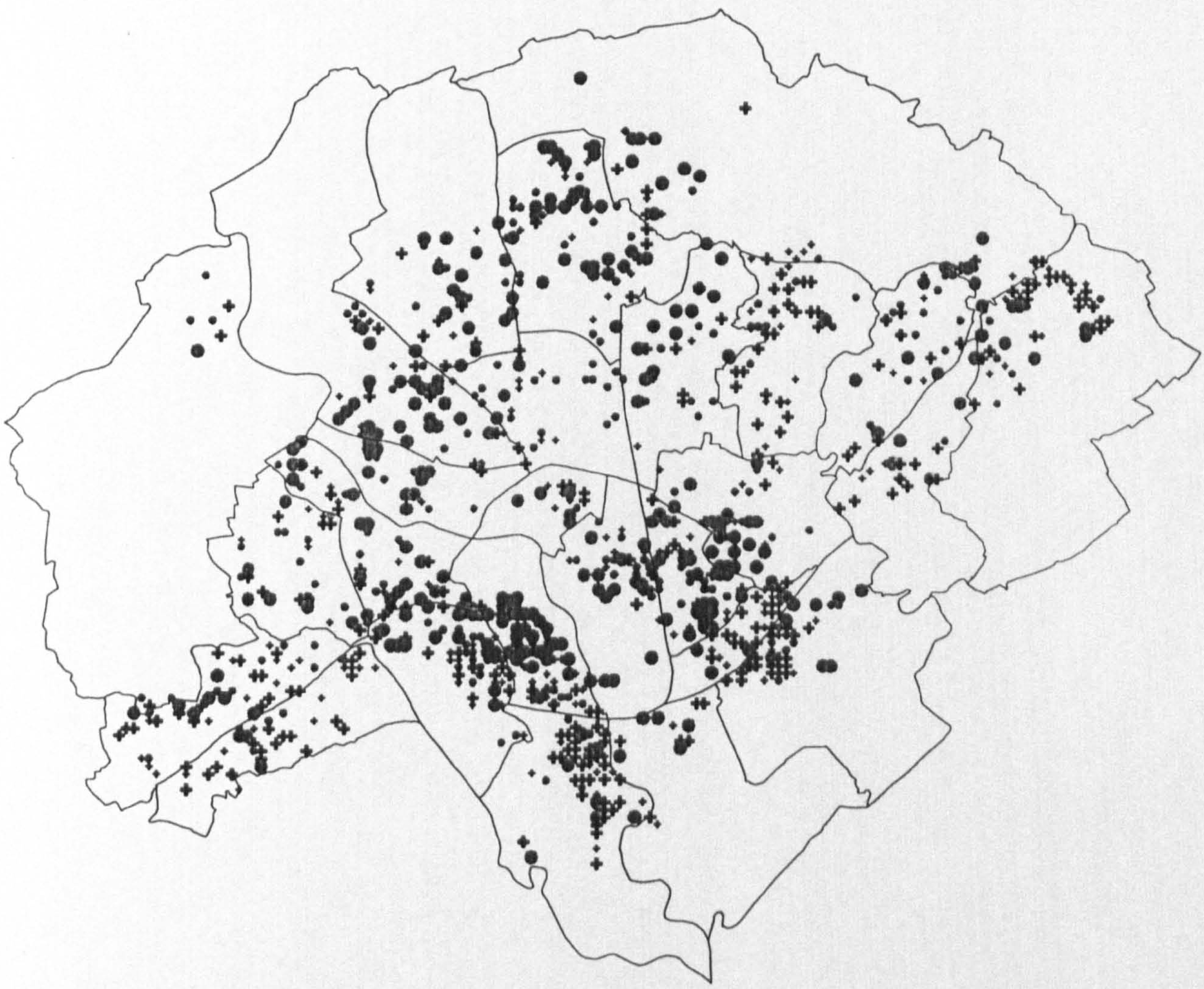


7.3.3.4 Conclusion

In an attempt to model the spatial dynamics of the Cardiff housing market, the fixed parameters of the traditional hedonic specification were expanded with respect to social class. The results of the subsequent models suggest that the housing market does not operate as a unified whole. Instead varying supply and demand schedules have caused submarket formation and spatial parameter drift of the structural attributes as hypothesized. More specifically, the results suggest that the housing market would appear to be segmented with respect to both housing bundles and location. As was concluded in the previous section, detached houses and bungalows appear to have separate market conditions. In addition, these market conditions vary across Cardiff, such that value of structural attributes drift with respect to socio-economic class. The only exception are bungalows, which tend to be located in Cyncoed and Rhiwbina, and hence do not vary across space, unlike the more ubiquitous detached housing bundles. The models also imply that the structural attributes drift in a non-linear fashion across space, such that they are cheaper in areas of either very high or very low social class. The corollary of this is that locational externalities become marginally more important for very affluent buyers, who are more likely to value location

Figure 7.10

Standard Residuals of Model 7.6



Residual Quartiles

- 0.613- 4.144
- 0 - 0.612
- -0.646 - 0
- -3.292 - -0.647

1 km

than average income buyers. Similarly, in areas of very low social class, housing in Local Authority are more desirable, possibly due to the marginally better locational externalities (lower density, more open space) than found in high density, inner city areas.

However, the spatial expansion specification is problematic. Diagnostic tests indicate that heteroscedasticity and spatial autocorrelation are still present, albeit in a reduced form. Furthermore, the residual maps reveal a strong community clustering of values, which remain substantially unchanged, despite the attempt to capture spatial variations. This distinct clustering of residuals into discrete areas highlights the main problem with the spatial expansion specification. It assumes that parameters drift in a continuous fashion across the housing market, whereas the underlying topology and other natural and artificial barriers are more likely to segment the housing market in a more descrete, contiguous fashion. It is also doubtful whether submarkets would operate at the level of the ED. Urban economic theory would appear to suggest that supply and demand schedules would only significantly differ across wider areas, such as communities. In addition, community context may also influence the effect of social class say, such that it would have a differential effect across the city. Finally, the specification assumes that the model variance is constant across the housing market, even though it may be argued that it may differ with respect to household income, and thus submarkets. These questions shall now be examined within the context of expanding the random parameters of the traditional hedonic specification.

Section 7.4 Spatially Expanding the Random Parameters: The Multi-level Specification

7.4.1 Introduction

The previous models were estimated at the level of the individual property, although the locational attributes were aggregated at the level of the ED. The multi-level specification requires the different spatial levels to be explicitly modelled: the individual property, the ED and the community, based upon Census wards. As has been explained in *Chapter Six*, the ED level captures the local variations that approximate individual streets, and are contextualised by the locational attributes. The community level approximates submarkets, at which housing attributes are deemed to vary. These three spatial levels were incorporated into the traditional hedonic specification by expanding the random parameters. Thus:

$$P_{ijk} = \alpha_{jk} X_{ijk} + \Sigma\, \beta_m Z_{mijk} + \varepsilon_{ijk} X_{ijk} \qquad \textbf{7.5}$$

where:
Z is the vector of m housing attributes,
k =1, ..., 26 is the number of communities,
j = 1, ..., 453 is the number of enumeration districts
i =1,1414 is the number of properties.

This is expanded using the equation:

$$\alpha_{jk} = \alpha + \mu\alpha_{jk} + \mu\alpha_k \qquad \textbf{7.6}$$

to form the random intercepts model:

$$P_{ijk} = \alpha_{jk} X_{ijk} + \Sigma\, \beta_m Z_{mijk} + (\, \mu\alpha_{jk} X_{ijk} + \mu\alpha_k X_{ijk} + \varepsilon_{ijk} X_{ijk}\,) \qquad \textbf{7.7}$$

As was discussed in *Chapter Two*, this specification has several conceptual and technical advantages over the previous two specifications. Firstly, it does not assume that housing market dynamics operate across Cardiff at a single spatial level of resolution. Instead, by expanding the random term, the specification can conceptualise the hierarchical nature of housing markets, in which houses are nested into streets, which in turn are nested into neighbourhoods. Moreover, this view of housing dynamics as operating across discrete space, such as neighbourhoods, is conceptually better than the traditional specification, in which the parameters are invariant, and the spatial parameter drift specification, in which the parameters vary smoothly and continuously, with no regard for the underlying urban morphology. The multi-level concept of house prices also distinguishes between the compositional effects of the housing stock (ie the structural attributes) and the contextual effects of the street and neighbourhood (ie the locational attributes). The previous specifications have treated structural and locational attributes as having an equal effect upon house price, even though the latter are typically shared amongst many properties. This leads to the problem of confounding the effects of structural attributes with locational attributes.

Table 7.9

Model 7.7 - The Grand Mean Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	56954	3449	16.50

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	4991	1485	3.36
ED Level			
CONSTANT	2562	338	7.59
Property Level			
CONSTANT	5739	257	22.30

-2*(log-likelihood) = 1142.35

To restate Wilkinson (1973): "[by using a single level, i]t is difficult conceptually as well as statistically to distinguish the effects of the characteristics of a dwelling alone on price since an obvious and important feature of neighbourhood is its stock of dwellings" (pp. 76). By taking into account the context of location, the multi-level specification will allow not only the structural attributes to vary across space, but also the locational attributes, ameliorating the problem of the spatial drift specification in which the effect of social class is constant across the city, regardless of neighbourhood context. Finally, unlike the previous specifications, the multi-level specification also has the added technical advantage of being able to handle inherently spatial data, resulting in the implicit modelling of spatial autocorrelation and spatial heteroscedasticity. The modelling was undertaken using the multi-level modelling package Mln (Rasbash et al, 1995).

7.4.2 Model 7.7 The Grand Mean Model

This is the simplest model, since it contains no fixed terms except for the overall intercept. This estimates the average house price for the whole of Cardiff, and gives a figure of £57,000 (Table 7.9), and compares to £56,000 for the ecological mean. The model allows the variation around this grand mean to be decomposed into variation at the level of the individual property, ED and community. The greatest variation occurs between individual houses within a community, although over a third of the variation occurs between communities. This is interesting since it suggests that the housing attribute prices vary significantly between places, which could be indicative of submarkets.

The significance of the parameters model may be judged by two methods (Woodhouse et al, 1996). The first is to calculate a t-statistic. This works well for fixed parameters, and is similar to significance testing in OLS regression. However, for random parameters, the distribution of the t-statistic may depart considerably from normality, especially in small samples. Instead, Woodhouse recommends using a likelihood ratio statistic. By checking if the likelihood ratios of successive models are significantly different, the significance of the additional terms in the model may be evaluated. Hence, the likelihood ratio statistic was calculated using the likelihood command in Mln. This will provide a base line against which the effects of including further terms may be judged.

Table 7.10 describes this variation in price around the grand mean. It is estimated that houses in Lisvane and St Mellons are some £40,000 more expensive than the average

Table 7.10

Community Level Premiums

Model 7.7 Community	Price	Model 7.8 Community	Price	Model 7.9 Community	Price
Lisvane & St Mellons	40700	Roath	16050	Roath	8759
Radyr & St Fagans	31158	Cyncoed	12958	Riverside	8223
Roath	30295	Llandaff	9058	Cyncoed	7228
Cyncoed	27296	Butetown	8490	Butetown	6620
Heath	20442	Heath	7985	Cathays	6364
Llandaff	19213	Cathays	7765	Plasnewydd	5782
Rhiwbina	18195	Riverside	7601	Llandaff	3857
Whitchurch & Tongwynlais	15114	Plasnewydd	6347	Heath	3263
Llanishen	14174	Canton	5268	Canton	2826
Canton	1003	Whitchurch & Tongwynlais	4580	Whitchurch & Tongwynlais	2722
Llandaff North	-1628	Rhiwbina	4224	Llanishen	1827
Rumney	-1714	Gabalfa	2114	Rhiwbina	1825
Gabalfa	-2700	Llanishen	1523	Llandaff North	1384
Plasnewydd	-4241	Radyr & St Fagans	152.7	Gabalfa	763.8
Butetown	-4267	Lisvane & St Mellons	-335.6	Lisvane & St Mellons	-233
Riverside	-5406	Llandaff North	-375	Adamsdown	-639
Cathays	-7627	Splott	-2381	Radyr & St Fagans	-1153
Trowbridge	-9977	Adamsdown	-2408	Splott	-1198
Fairwater	-10349	Grangetown	-2762	Grangetown	-1923
Grangetown	-12060	Rumney	-4043	Fairwater	-3250
Ely	-14567	Fairwater	-4779	Rumney	-3397
Pentwyn	-14871	Pentwyn	-9904	Trowbridge	-7141
Caerau	-15497	Trowbridge	-11313	Llanrumney	-7614
Llanrumney	-15812	Caerau	-11850	Ely	-8504
Splott	-16982	Llanrumney	-12285	Caerau	-8829
Adamsdown	-19176	Ely	-13143	Pentwyn	-9212

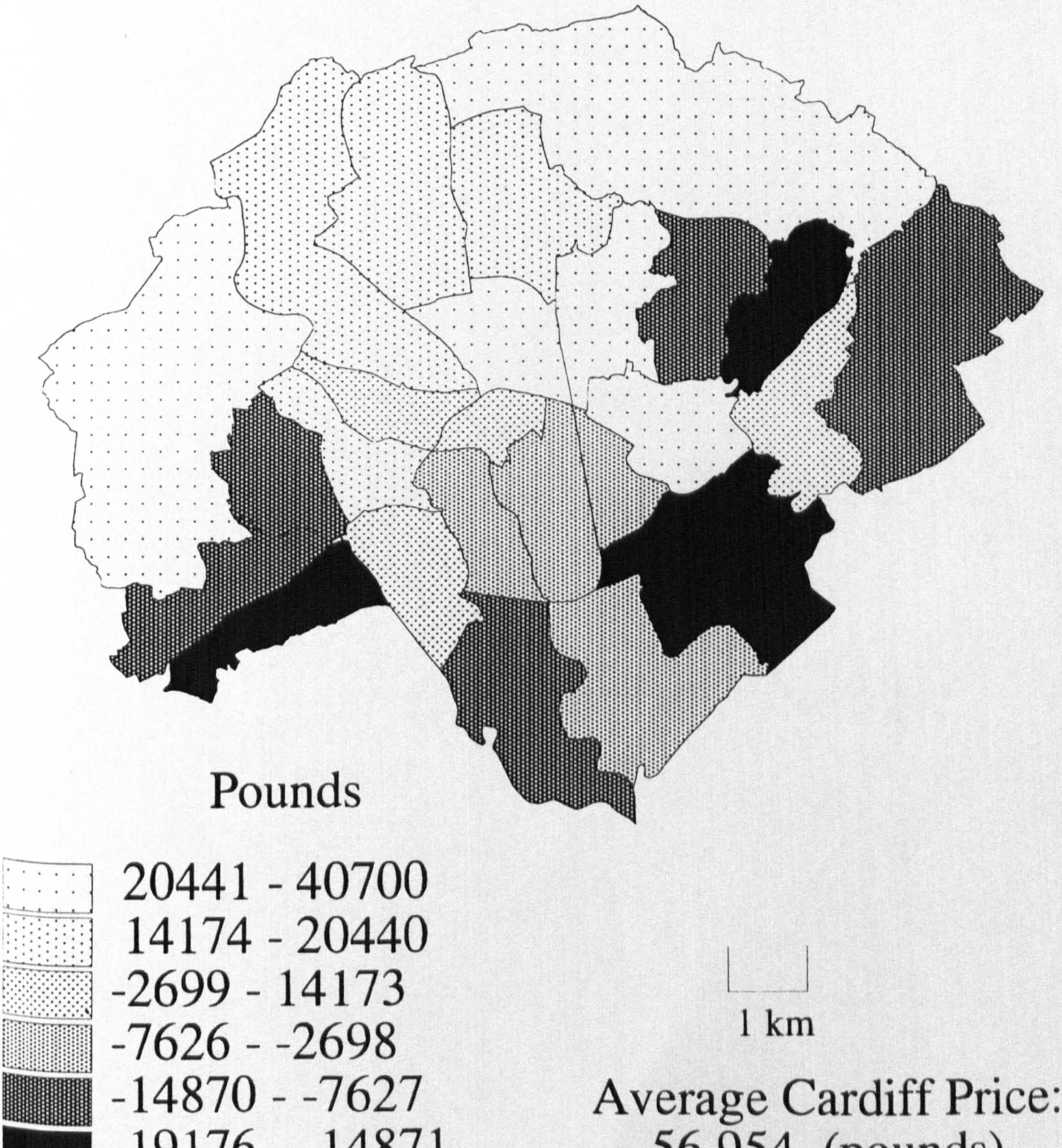
Table 7.10 (cont)

Community Level Premiums

Model 7.10 Community	Price	Model 7.11 Community	Price	Model 7.12 Community	Price
Roath	8314	Riverside	10481	Riverside	9982
Riverside	7723	Roath	8706	Roath	7913
Cyncoed	7019	Cyncoed	5471	Cyncoed	5749
Butetown	6131	Llandaff	4865	Plasnewydd	4575
Cathays	5899	Plasnewydd	4681	Llandaff	4468
Plasnewydd	5420	Cathays	3836	Cathays	4025
Llandaff	4332	Heath	3350	Whitchurch & Tongwynlais	3331
Heath	3657	Canton	3270	Heath	3252
Whitchurch & Tongwynlais	3058	Whitchurch & Tongwynlais	3037	Llanishen	2835
Canton	2855	Rhiwbina	2591	Rhiwbina	2684
Llanishen	2486	Butetown	2373	Canton	2611
Rhiwbina	2334	Llanishen	2248	Llandaff North	1946
Llandaff North	1363	Llandaff North	1758	Butetown	1864
Lisvane & St Mellons	809	Lisvane & St Mellons	811	Lisvane & St Mellons	1325
Gabalfa	531	Gabalfa	445	Radyr & St Fagans	511
Radyr & St Fagans	-376	Radyr & St Fagans	437	Gabalfa	233
Grangetown	-2160	Grangetown	-1695	Grangetown	-2098
Adamsdown	-2206	Fairwater	-2646	Fairwater	-2599
Splott	-2662	Rumney	-2772	Rumney	-2761
Fairwater	-3041	Splott	-3673	Splott	-3804
Rumney	-3250	Adamsdown	-4552	Adamsdown	-4030
Trowbridge	-6864	Llanrumney	-5999	Llanrumney	-5857
Llanrumney	-7257	Trowbridge	-6210	Trowbridge	-5953
Ely	-8302	Ely	-6406	Ely	-6191
Caerau	-8624	Caerau	-8407	Caerau	-8446
Pentwyn	-9233	Pentwyn	-8818	Pentwyn	-8789

Figure 7.11

Differential Community Prices: Model 7.7



Cardiff price, whilst houses in Adamsdown were nearly £20,000 cheaper. This is illustrated graphically in Figure 7.11. Here, the cheapest communities are concentrated within the Inner Area, and peripheral estates with the highest premiums found in the northern suburbs. As Jones & Bullen (1994) have commented, the model has merely estimated a grand average for Cardiff, and individual averages for each community, and has therefore reproduced ecological estimates, similar to those in *Chapter Six*. However, these are precision-weighted, so that the estimate for the community effect based on a small number of sales are down-weighted to the overall average price for Cardiff.

7.4.3 Model 7.8 - The Structural Attributes Model

The structural attributes model includes all the level-1 housing attributes, together with distance to the city centre, and will allow an assessment of the contextual effects of location after adjusting for the compositional effects of the housing stock (Table 7.11). The constant now represents the average price of the typical property, which is estimated at £45000, reflecting the price of an average sized terraced house. There are several interesting features, notably the insignificance of semi-detached floor area, and the negative sign for bungalows with two bathrooms. The latter has previously been discussed is terms of spatial parameter drift and can be regarded as a mis-specification caused by omitted social class interactions. Distance to the CBD also has a counter-intuitive sign, although again this is a reflection of omitted higher level interactions.

With these exceptions, the fixed term estimates are as expected and are comparable to the previous models, but their inclusion has had an interesting effect on the random terms. The inclusion of the structural attributes has resulted in a decline of the property level variance, an obvious result since price differences between individual houses are a result of differences in structural attributes. Nearly half of all the variation now occurs between communities, a substantial increase from the previous model. An examination of the community-level differences shows that there are major changes in both the rank of communities, and also in the size of their contextual effects. The previous most expensive communities, Lisvane & St Mellons and Radyr and St Fagans, are now average for Cardiff, whilst previously below average community such as Butetown and Cathays, are now substantially above average. Also, the size of the community premiums have declined substantially, suggesting that they were capturing the compositional effects of the housing stock. In terms of structural attributes, buyers are getting much less for their money in areas

Table 7.11

Model 7.8 - The Structural Attributes Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	44801	2321	19.30
Floor AREA	36.11	1.69	21.43
FLOOR SD	-5.59	5.29	-1.07
FLOOR D	3.56	0.28	12.57
FLOOR B	2.66	0.39	6.71
D BATH 2	5340	768	6.95
B BATH 2	-1532	615	-2.49
D SHOWER 1	2076	445	4.66
FULL CH	4092	924	4.43
GARAGE	5651	658	8.59
ORP	4244	804	5.28
GDN:NONE	-4650	2114	-2.20
GDN:5-50M	4443	859	5.17
GDN:>50M	7506	959	7.83
NEEDS MODS	-5782	2149	-2.69
DIST CBD	1.92	0.49	3.93

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	1553	451	3.44
ED Level			
CONSTANT	487	67	7.21
Property Level			
CONSTANT	1239	54	22.74

-2*(log-likelihood) = -661.07

Table 7.12

Model 7.9 - The Full Housing Attributes Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	44801	2321	19.30
FLOOR AREA	36.11	1.69	21.43
FLOOR SD	-5.59	5.29	-1.07
FLOOR D	3.56	0.28	12.57
FLOOR B	2.66	0.39	6.71
D BATH 2	5340	768	6.95
B BATH 2	-1532	615	-2.49
D SHOWER 1	2076	445	4.66
FULL CH	4092	924	4.43
GARAGE	5651	658	8.59
ORP	4244	804	5.28
GDN:NONE	-4650	2114	-2.20
GDN:5-50M	4443	859	5.17
GDN:>50M	7506	959	7.83
NEEDS MODS	-5782	2149	-2.69
DIST CBD	1.92	0.49	3.93
SOCIAL	3057	265.40	11.52
LA > 50%	-14798	3281	-4.51
SOCLA	-4635	907	-5.11

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	702	209	3.36
ED Level			
CONSTANT	255	50	5.08
Property Level			
CONSTANT	1239	54	22.74

-2*(log-likelihood) = -815.68

like Roath and Cyncoed, than in areas like Llanrumney and Ely. Instead they are spending more money upon location, a conclusion also suggested by the results of the spatial parameter drift models. The difference in the likelihood ratio statistic of this model and the Grand Mean model is 1803.42, which under the null hypothesis follows a chi-squared distribution with degrees of freedom equal to the number of new parameters, in this case 15 (Woodhouse et al, 1996). The probability of obtaining a chi-square of this magnitude by chance is exceedingly small (less than 0.001), strongly indicating that the structural attributes have an important effect in explaining house price variation in the model.

7.4.4 Model 7.9 The Full Housing Attributes Model

The full housing attribute model has the same random and property level fixed terms as the previous model, but now includes the locational attributes at the ED level (Table 7.12). Since these do not vary at level-1, the fixed and random estimates for the property level attributes remain unchanged. However, the ED level and community level random effects have been reduced, resulting in the property level explaining over a half of house price variations. The variable measuring housing quality was insignificant in the model (and hence was omitted), whilst socio-economic class had a significant effect upon house price differentials as was expected. These represent the relationship between house price and locational attributes after the compositional effects of the structural attributes have been allowed for. The addition of locational attributes at the ED level has resulted in marked changes at the community level. Firstly, the effects of area are now smaller, and on average, the premiums have halved for most communities. Furthermore, there has been some interesting changes in rank, notably the promotion of Riverside and Llandaff North. This suggests that these areas command a higher premium, given the social class of the areas, and may be caused by unaccounted externalities, in this case proximity to Bute Park and Llandaff Cathedral respectively. The cheapest communities are those characterised by a high proportion of old Local Authority stock and privately rented properties, implying a stigma effect. Again, the difference in the likelihood ratios (154.61) indicates that the addition of the locational attributes have had a significant effect upon explaining the variation in the model.

7.4.5 Higher Level Interactions

The above three models assume that the structural attributes are constant across Cardiff, and that the areal differences can be captured in a single variance term. An equivalent single level model would be similar to the traditional specification, except that the constant term would be expanded to accommodate community dummy variables. In such a specification, no structural parameter drift would occur, but instead there would be an additional premium to locate in a particular community. However, as has previously been discussed, such a specification would be inaccurate. Disequilibrium in supply and demand presuppose segmentation of the housing market, and this may lead to submarket formation, and hence the spatial variation of attribute prices. This was identified in the results from the spatial drift specifications, which indicated that the implicit prices of certain attributes appear to drift with respect to social class. In Model 7.9, a third of the house price variation occurs between communities after compositional effects of the housing stock and contextual effects of location have been taken into account. These unexplained community differentials may be caused by variation in structural parameters at the community level. Therefore, the nature of structural attribute variation across space and the possible presence of submarket shall now be examined.

Previous attempts to measure the effects of submarkets have in the main relied upon the switching regression technique (see *Chapter Two*), in which an hedonic model is estimated for the entire market and then separately for each individual community. Significant differences between community models, and the overall model are indicative of submarkets. However, it has previously been explained that this method has several problems, a principal one being that by estimating each model separately, no overall housing market dynamics are captured across the wider city. This can be ameliorated by using the multi-level specification, since all the data is used when estimating the structural attribute prices for each community. In this case, if the structural attributes are allowed to vary at the community level, any significant difference in the resulting random terms will be indicative of possible submarkets

Therefore, the random intercepts model (equation 7.7) was expanded with respect to equation 7.8:

$$\beta_{mjk} = \beta_m + \mu\beta_{mk} \qquad \textbf{7.8}$$

Table 7.13

Model 7.10
Floor Area - Community Interactions Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	45705	1722	26.53
FLOOR AREA	39.11	3.94	9.92
FLOOR D	9.10	1.04	8.71
D BATH 2	4671	214	6.83
B BATH 2	-1532	615	-2.49
D SHOWER 1	2778	154	5.05
FULL CH	4524	854	5.29
GARAGE	5843	707	8.26
ORP	3510	880	3.98
GDN:NONE	-4631	1759	-2.63
GDN:5-50M	4050	844	4.80
GDN:>50M	6790	917	7.41
NEEDS MODS	-5673	2033	-2.79
DIST CBD	-0.9691	0.41	-2.35
SOCIAL	3103	316	9.83
LA > 50%	-15843	5835	-2.72
SOCLA	-4868	1160	-4.20

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	680	204	3.33
FLOOR AREA	0.22	0.07	3.12
FLOOR AREA / CONSTANT	0.13	0.17	0.79
ED Level			
CONSTANT	269	52	5.22
Property Level			
CONSTANT	1266	57	22.40

-2*(log-likelihood) = -837.46

To form the fully random model:

$$P_{ijk} = \alpha_{jk} X_{ijk} + \Sigma \beta_m Z_{mijk} + (\mu\alpha_{jk} X_{ijk} + \mu\alpha_k X_{ijk} + \mu\beta_{mk} Z_{mijk} + \varepsilon_{ijk} X_{ijk}) \qquad 7.9$$

In which housing attribute Z_m is allowed to vary at the community level

7.4.5.1 Model 7.10 - Floor Area Interactions

Since floor area is the main structural attribute, this was allowed to vary at the community level. The random term in Table 7.13 measures the variation in the price of floor area between communities whilst the covariance term measures the relationship between average community house price and the price of floor area. The addition of these random terms has caused a difference of 21.78, which is significant with 2 degrees of freedom. However, the standard error of the covariance term is rather large relative to the coefficient, suggesting that this term may have an insignificant effect upon explaining house price variation. This was verified by removing the covariance term, and re-estimating the model. There was negligible difference with the new parameter estimates whilst the new likelihood ratio (-864.58) indicated that the removal of the covariance term has not had little effect upon the model. Hence, the model suggests that the price of floor area does indeed vary between communities, whilst the insignificance of the covariance term suggests that there is no relationship between average community house price and the price of floor area. In other words, it is not the case that the increase in the price of floor area is greater for expensive properties than cheaper ones or vice versa.

Table 7.14 summarises the implicit prices of floor area for Cardiff and each community. On average across Cardiff, an extra square foot of space would be worth £38.34, whilst this varies from place to place. In Cyncoed, an extra square foot of space would be valued at £47.83, whilst in Adamsdown it would only be £26.84. These differences reflect the differences in the supply and demand schedules that operate in these communities. Figure 7.12 shows the change that the addition of the random floor area term has had on the implicit price of community. It is clear that there is a Inner Area/suburban split, with suburban communities becoming more expensive once the differential price of floor area has been taken into account and vice versa. However, the size of this effect is small, and there has been little change in rank.

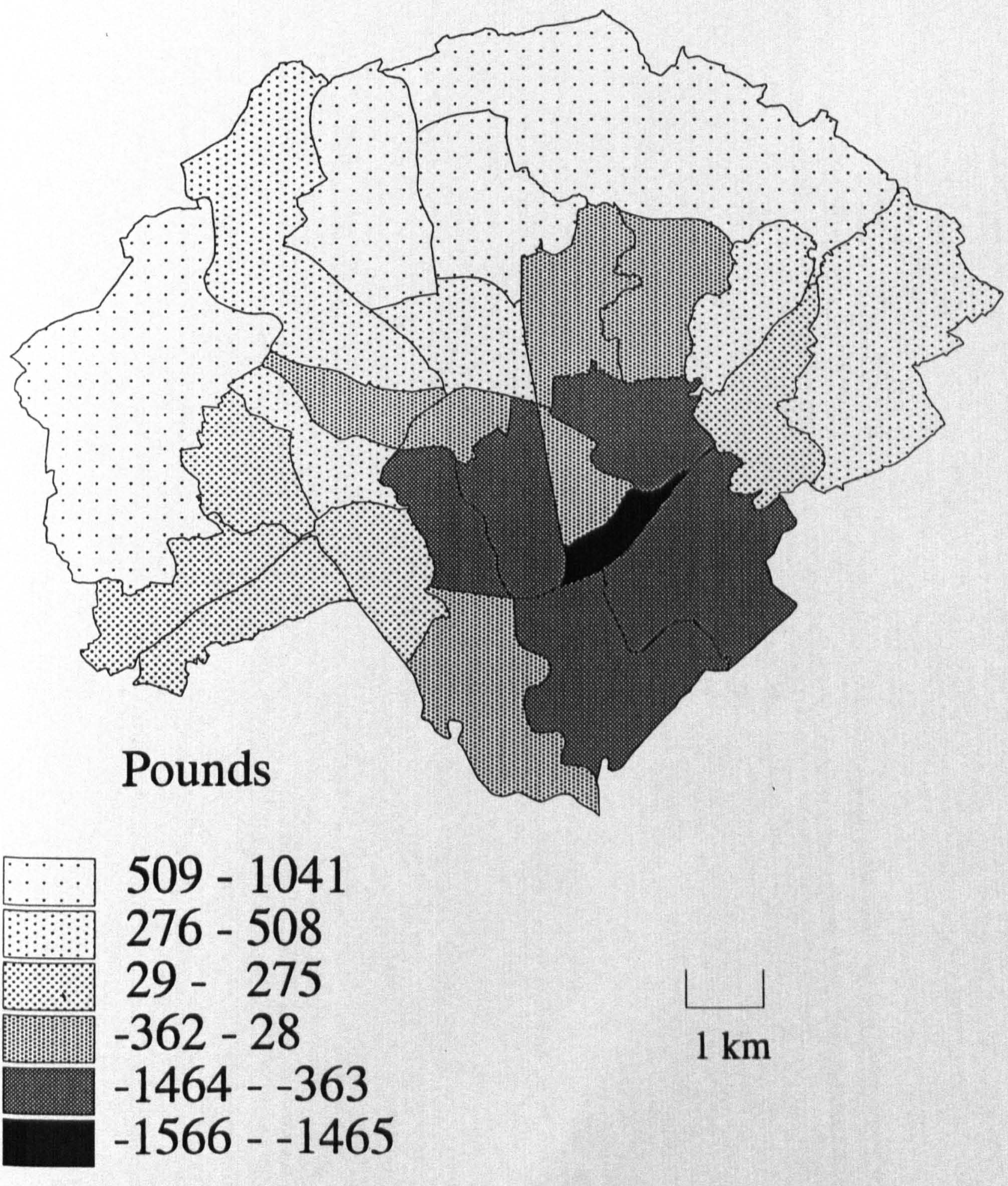
Table 7.14

Community Level Floor Area Differential Prices

Model 7.10 Community	Price	Model 7.11 Community	Price	Model 7.12 Community	Price
Cyncoed	9.49	Llanishen	9.20	Llanishen	9.11
Llanishen	8.71	Cyncoed	9.16	Cyncoed	8.88
Whitchurch & Tongwynlais	6.27	Whitchurch & Tongwynlais	6.52	Whitchurch & Tongwynlais	6.24
Llandaff	4.37	Llandaff	4.78	Llandaff	4.47
Roath	3.95	Roath	4.39	Roath	4.26
Rhiwbina	3.21	Rhiwbina	3.50	Rhiwbina	3.70
Plasnewydd	2.92	Llanrumney	3.14	Plasnewydd	3.03
Llanrumney	2.43	Plasnewydd	3.04	Llanrumney	2.74
Heath	1.90	Heath	2.53	Heath	2.07
Caerau	0.48	Lisvane & St Mellons	0.84	Fairwater	0.80
Cathays	0.44	Fairwater	0.43	Lisvane & St Mellons	0.62
Lisvane & St Mellons	0.06	Radyr & St Fagans	0.13	Radyr & St Fagans	0.15
Fairwater	0.04	Caerau	0.07	Riverside	-0.12
Butetown	-0.01	Llandaff North	-0.07	Caerau	-0.14
Trowbridge	-0.30	Riverside	-0.14	Cathays	-0.14
Radyr & St Fagans	-0.34	Ely	-0.47	Llandaff North	-0.23
Llandaff North	-0.53	Cathays	-1.15	Butetown	-0.37
Riverside	-0.63	Pentwyn	-1.25	Ely	-1.48
Ely	-1.39	Trowbridge	-2.17	Pentwyn	-1.80
Pentwyn	-1.51	Gabalfa	-3.02	Trowbridge	-1.96
Gabalfa	-3.01	Canton	-3.19	Gabalfa	-2.78
Rumney	-3.18	Rumney	-3.27	Canton	-3.08
Canton	-3.75	Butetown	-3.60	Rumney	-3.27
Grangetown	-7.25	Grangetown	-6.44	Grangetown	-7.53
Splott	-10.80	Splott	-10.8	Splott	-11.40
Adamsdown	-11.50	Adamsdown	-12.10	Adamsdown	-11.80

Figure 7.12

Changes in Community Prices due to Level 3 Floor Area Interactions



The inclusion of the floor area random term has had a small, but important effect upon the attribute prices at the individual house level. Firstly, the variable measuring bungalow floor area has became insignificant. This implies this was probably capturing the differential floor area price now accounted for by the higher level random terms. More specifically, it would appear that the variable was capturing the contextual effects operating upon floor area in Cyncoed and Rhiwbina, the location of the majority of bungalows. Secondly, the measure of accessibility to the CBD has become negative and significant. Again, this would imply that the variable had captured the contextual effects of community upon floor area. Table 7.14 illustrates that floor area in suburban communities are generally more expensive than Inner Area ones, and hence the counter-intuitive positive relationship between distance and house price. The price of floor area for detached properties does not vary significantly between communities, but is uniformly more expensive than all other properties across the city as a whole. The remaining structural attributes remain unchanged, and none of them varied significantly at the community level.

7.4.5.2 Model 7.11 Social Class - Between Community Interactions

Previous empirical research has suggested that social class is an important factor in house price variation, and this was tested using the spatial drift specification, which established that the implicit price of certain attributes, such as floor area, interacted with socio-economic class. Model 7.11 captures this effect by allowing social-economic class to vary at the community level. This model also allows the effect of social class to vary depending upon community context, a concept that was not possible in the spatial drift specification, which assumed a constant effect. Table 7.15 shows that, in terms of the fixed attribute prices, the counter-intuitive price differential between single and two bathroom bungalows bungalows becomes insignificant and falls out of the model, indicating that it was capturing a social class effect as was suggested in the previous section

The significance of the three additional random terms were evaluated by the size of their standard errors, and the likelihood ratio statistic. These indicated that the random term for social class had an insignificant effect upon the model, whilst the floor area / constant covariance term, which was previously insignificant, now had a significant effect. Therefore, although the effect social class *per se* does not vary significantly between communities, it does interact with floor area and average house price at this higher level, such that the marginal price / floor area relationship is steeper in areas of higher social class,

Table 7.15

Model 7.11
Social Class - Between Community Interactions Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	45252	1639	27.61
FLOOR AREA	38.34	3.955	9.69
FLOOR D	8.57	1.031	8.31
D BATH 2	4511	2136	6.79
D SHOWER 1	2763	1529	5.19
FULL CH	4414	831	5.31
GARAGE	5734	727	7.88
ORP	3470	827	4.20
GDN:NONE	-4955	1772	-2.79
GDN:5-50M	3843	834	4.61
GDN:>50M	6556	899	7.29
NEEDS MODS	-5556	2069	-2.69
DIST CBD	-0.7021	0.422	-1.67
SOCIAL	2983	368	8.11
LA > 50%	-10161	3461	-2.94
SOCLA	-3644	1025	-3.56

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	633	198	3.20
FLOOR AREA	0.00089	0.00039	2.25
SOCIAL	30.65	15.88	1.93
FLOOR AREA / CONSTANT	0.25	0.083	3.05
FLOOR AREA / SOCIAL	0.053	0.019	2.74
CONSTANT / SOCIAL	15.79	39.47	0.40
ED Level			
CONSTANT	215	48	4.53
Property Level			
CONSTANT	1246	56	22.45

-2*(log-likelihood) = -865.76

Figure 7.13
Price / Floor Area Relationship With
Social Class Interactions: Model 7.11

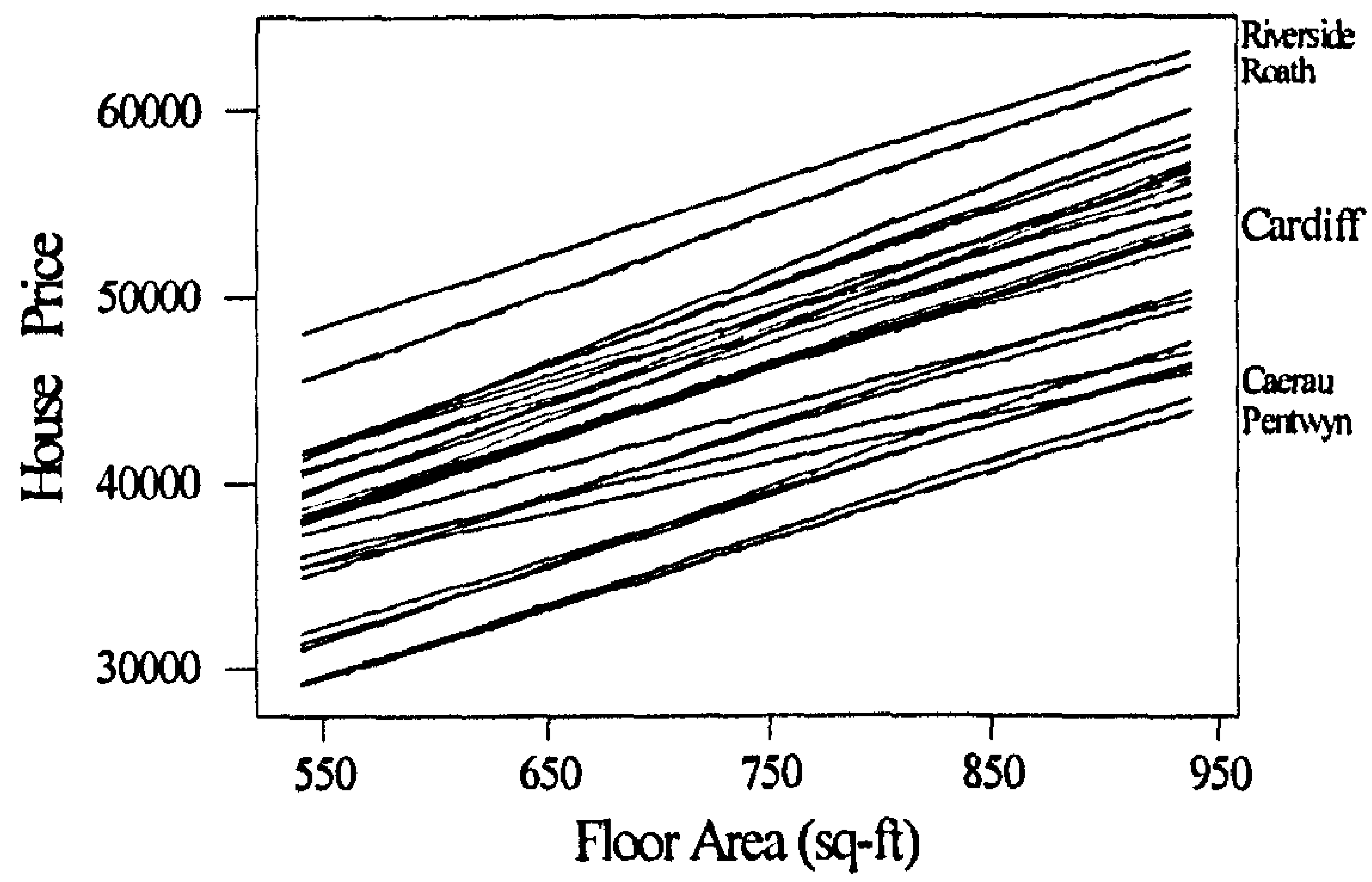
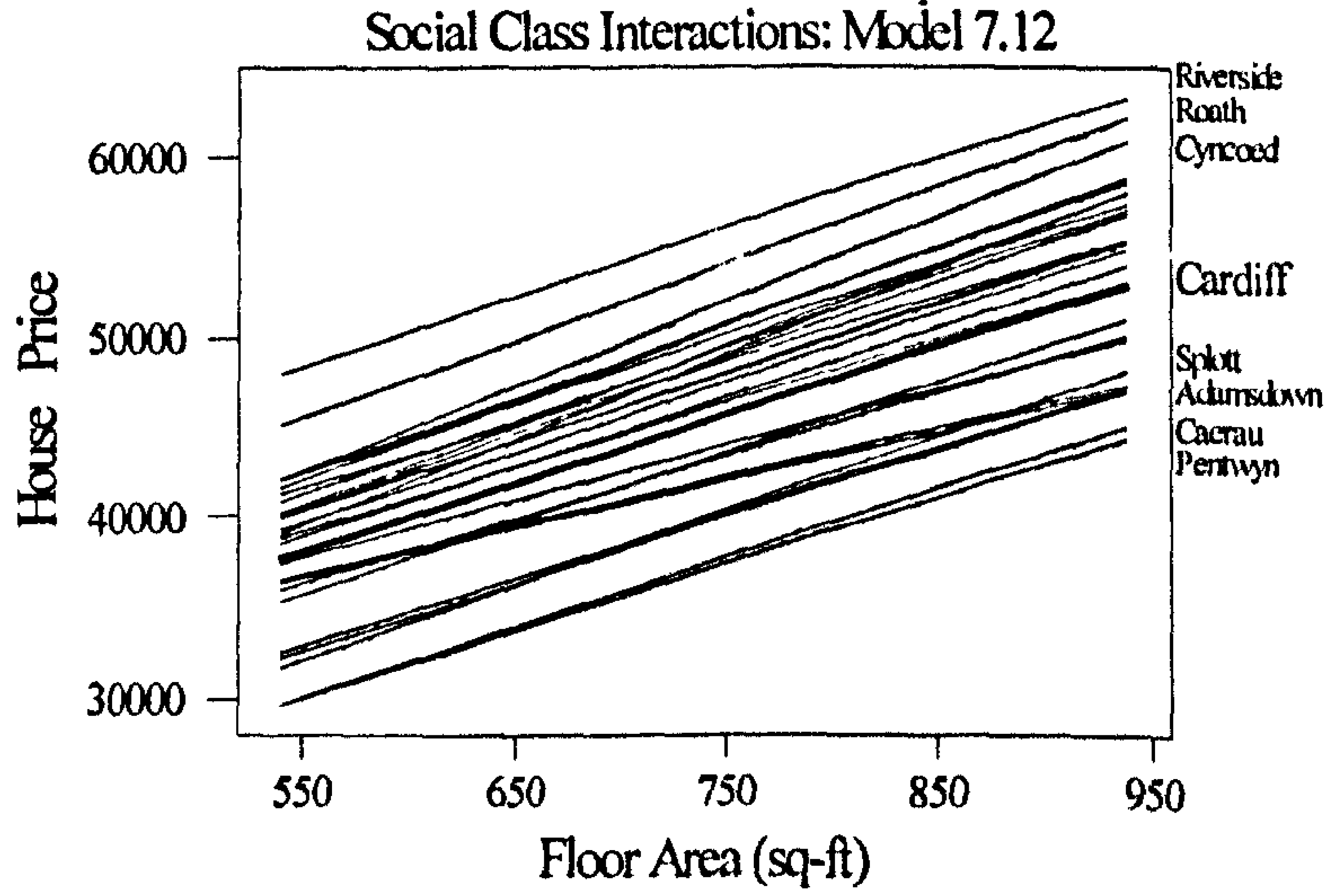


Figure 7.14
Price / Floor Area Relationship With
Social Class Interactions: Model 7.12



whilst average house prices are more expensive. As a consequence, the covariance between floor area and typical community house price has now become significant, with Figure 7.13 implying that the differences in community property prices are generally smaller for larger houses than smaller houses. These results suggest that the difference a community makes depends upon the size of the property, and the social class of the area. An examination of the community effects (Table 7.10) indicates that the most expensive communities are those with a combination of large properties and high social class, and vice-versa. Areas of high social class and (relatively) small properties, such as Lisvane and St Mellons, or large properties and average social class, such as Radyr & St Fagans and Rhiwbina, only have an average premiums.

7.4.5.3 Model 7.12 Social Class - Within Community Interactions

The above model has allowed social class to vary at the community level. However, the spatial parameter drift models showed that social class operates in a non-linear fashion, such that the greatest effects occur in areas of either very high or very low social class. This was specified through the use of a quadratic functional form. In the multi-level specification, however, such an effect can be achieved by allowing the effect of social class to vary within communities, at the level of the ED. Coupled with variations at the community level, such a model will allow a complex geography of social class interactions. Table 7.16 summarises the results of this model. The influence of these ED level random terms are substantial (differences in likelihood of 18.12 with 2 degrees of freedom). The constant term becomes insignificant, implying that within a community, average house price does not vary significantly at the ED level, whilst the significant social class random term accounts for the non-linear effect of social class.

At the community level, the significance of the random terms become greater (since their standard errors decrease), although the random term for social class still remains insignificant. An examination of the changes in the floor area differentials suggests that those communities with the greatest mix of social class have experienced the greatest changes. For instance, Butetown, which is traditionally a low social class area but now has enclaves of high social class in the redeveloped docklands, has experienced an increase in floor area of £3.23 per square foot. Conversely, in neighbouring Grangetown, the floor area differential has declined in price by £1.10 per square foot, reflecting the higher concentrations of lower class areas. Figure 7.14 describes the relationship between average

Table 7.16

Model 7.12
Social Class - Within Community Interactions Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	45707	1645	27.79
FLOOR AREA	38.62	3.93	9.84
FLOOR D	9.53	0.75	12.64
D BATH 2	4206	619	6.79
D SHOWER 1	2553	492	5.19
FULL CH	4460	779	5.73
GARAGE	5822	639	9.11
ORP	3267	687	4.76
GDN:NONE	-4974	1647	-3.02
GDN:5-50M	3921	862	4.55
GDN:>50M	6643	1021	6.51
NEEDS MODS	-5529	1589	-3.48
DIST CBD	-0.10	0.34	-2.95
SOCIAL	3333	283	11.77
LA > 50%	-11350	4687	-2.42
SOCLA	-3606	1347	-2.68

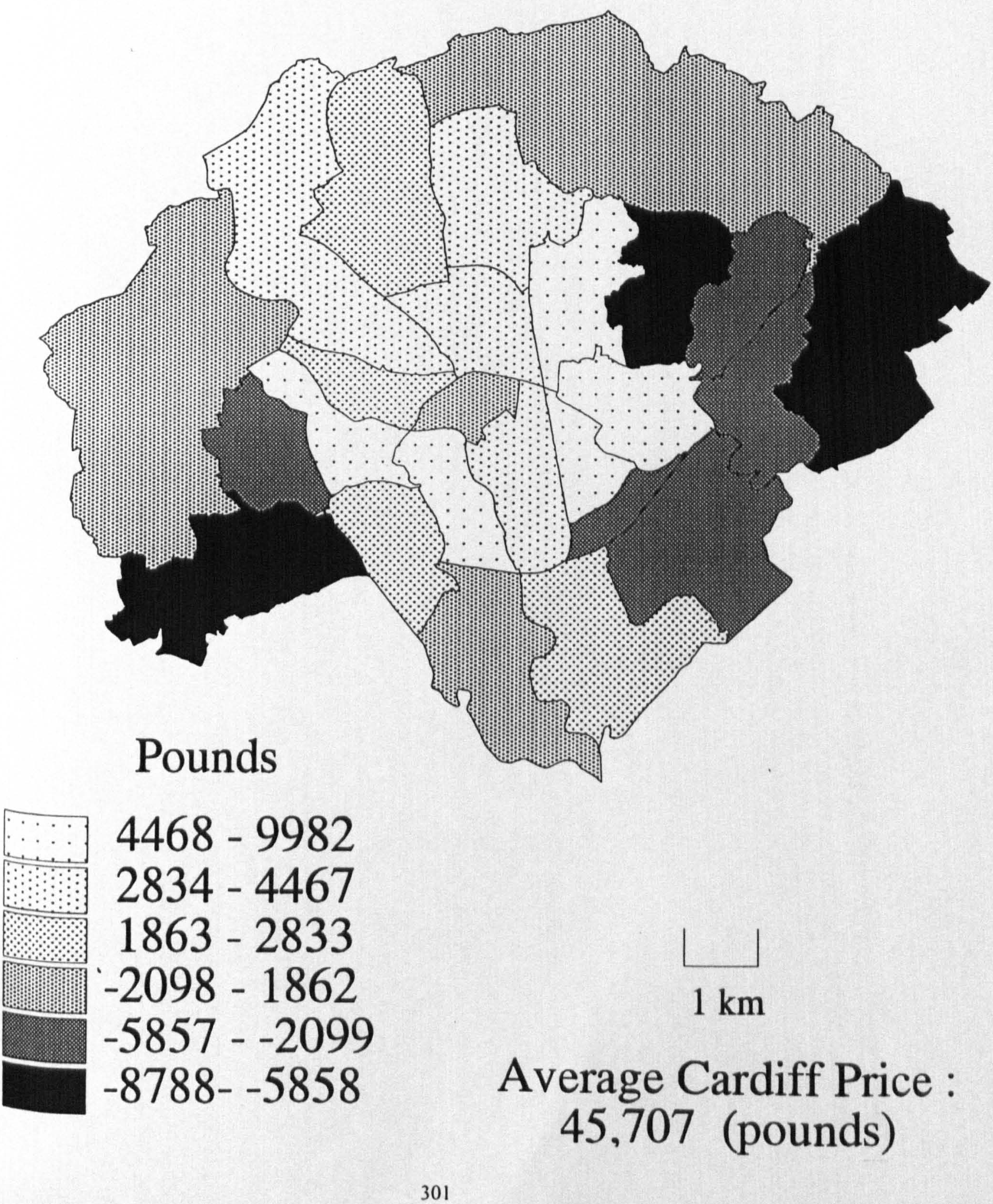
RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	588	177	3.33
FLOOR AREA	0.00087	0.00023	3.78
SOCIAL	14.17	11.24	1.26
FLOOR AREA / CONSTANT	0.27	0.084	3.26
FLOOR AREA / SOCIAL	0.025	0.0088	2.87
CONSTANT / SOCIAL	21.32	7.84	2.72
ED Level			
CONSTANT	90	51	1.77
SOCIAL	44.27	15.70	2.82
CONSTANT / SOCIAL	7.30	15.22	0.48
Property Level			
CONSTANT	1245	56	22.37

-2*(log-likelihood) = -883.88

Figure 7.15

Differential Community Prices: Model 7.15



community house price and the price of floor area. The effect of allowing social class to vary within a community has increased the strength of this relationship in the more expensive communities, such as Cyncoed, whilst having very little effect on communities of below average price. Two communities that appear to have a significantly different relationship from the rest of Cardiff are Splott and Adamsdown, which have much gentle price gradients. Since these are adjacent communities, the supply and demand mechanisms in these communities maybe very similar, compared to elsewhere.

Figure 7.15 illustrates the geography of the differential community prices. The most expensive communities tend to be located either in, or adjacent to the Inner Area, which contrasts with Figure 7.11, in which the most expensive were on the edge of the suburbs. The cheapest communities correspond to the peripheral Local Authority estates, corroborating the evidence of the 'stigma effect' attached to these areas. This gives an idea of which communities require additional premiums, and could be viewed as a potential method of limiting accessibility to these areas.

7.4.6 The Relationship Between Social Class and Floor Area

In the previous section, it was argued that in areas of higher social class, buyers spend marginally more on locational attributes than on structural attributes, and this is reflected in the decreasing price per unit floor area. This assertion can be re-evaluated using the results from the multi-level models. Figure 7.16 is the result of plotting the unit price of floor area in each community (Model 7.12) against community social class. With the exception of Butetown, it can be seen that the price of floor area initially increases, before decreasing in areas of high social class. Butetown is an anomaly, caused by the Docklands development in the an area traditionally of low social class.

Figure 7.17 describes the relationship between the price per unit floor area and average community house price differentials. This summarises the previous finding of a positive linear relationship between the price of floor area and the average value of a house in a particular community. If it is assumed that the most expensive structural attributes will be located in those communities with the most expensive locational attributes, as the graph predicts, then this can be used to evaluate the relationship between the value of housing in terms of its structural and locational attributes. For instance, in Roath, Cyncoed and Rhiwbina, buyers are spending marginally more on both the physical housing stock and

Figure 7.16
Price per unit Floor Area by Community Level
Social Class: Model 7.12

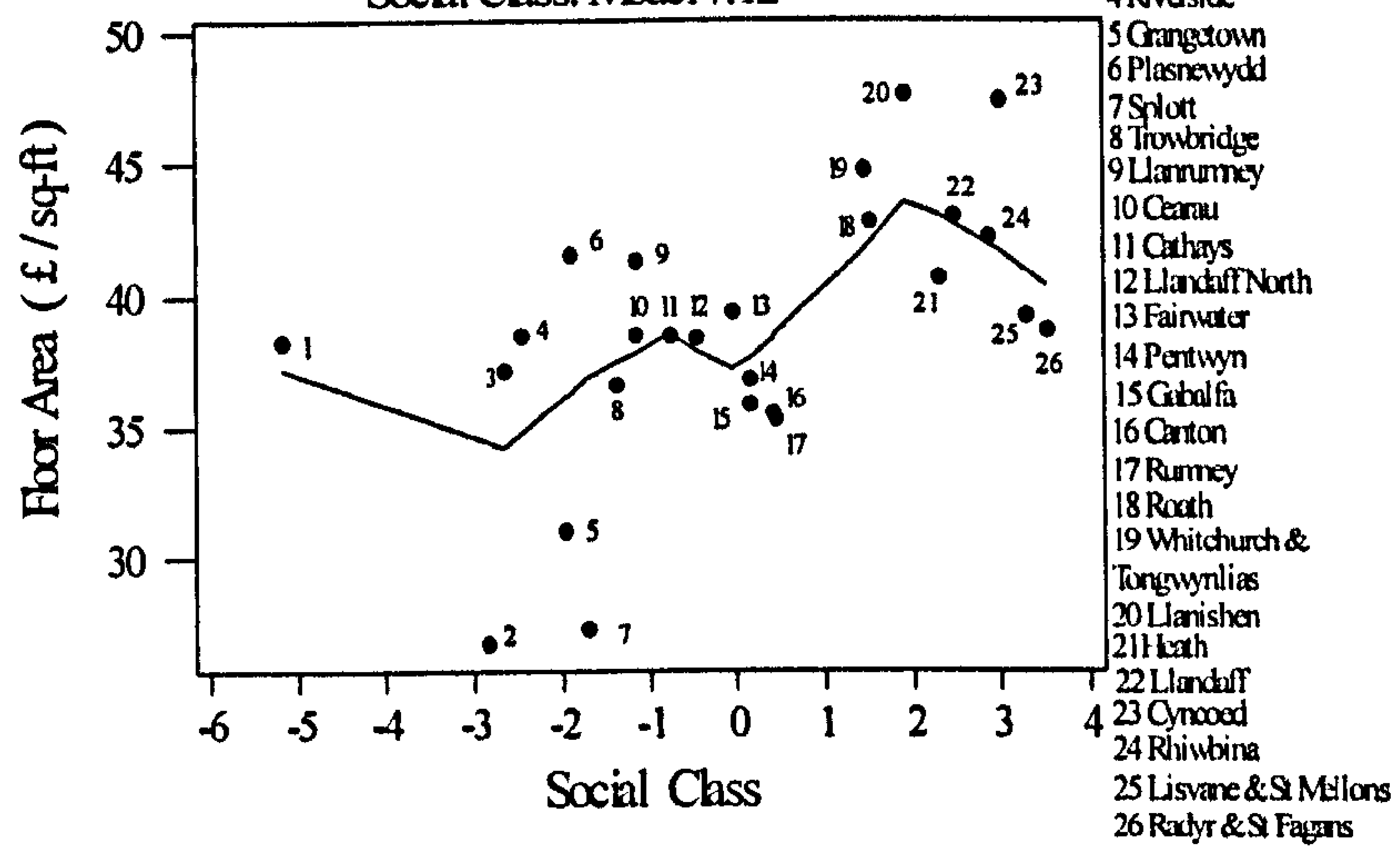
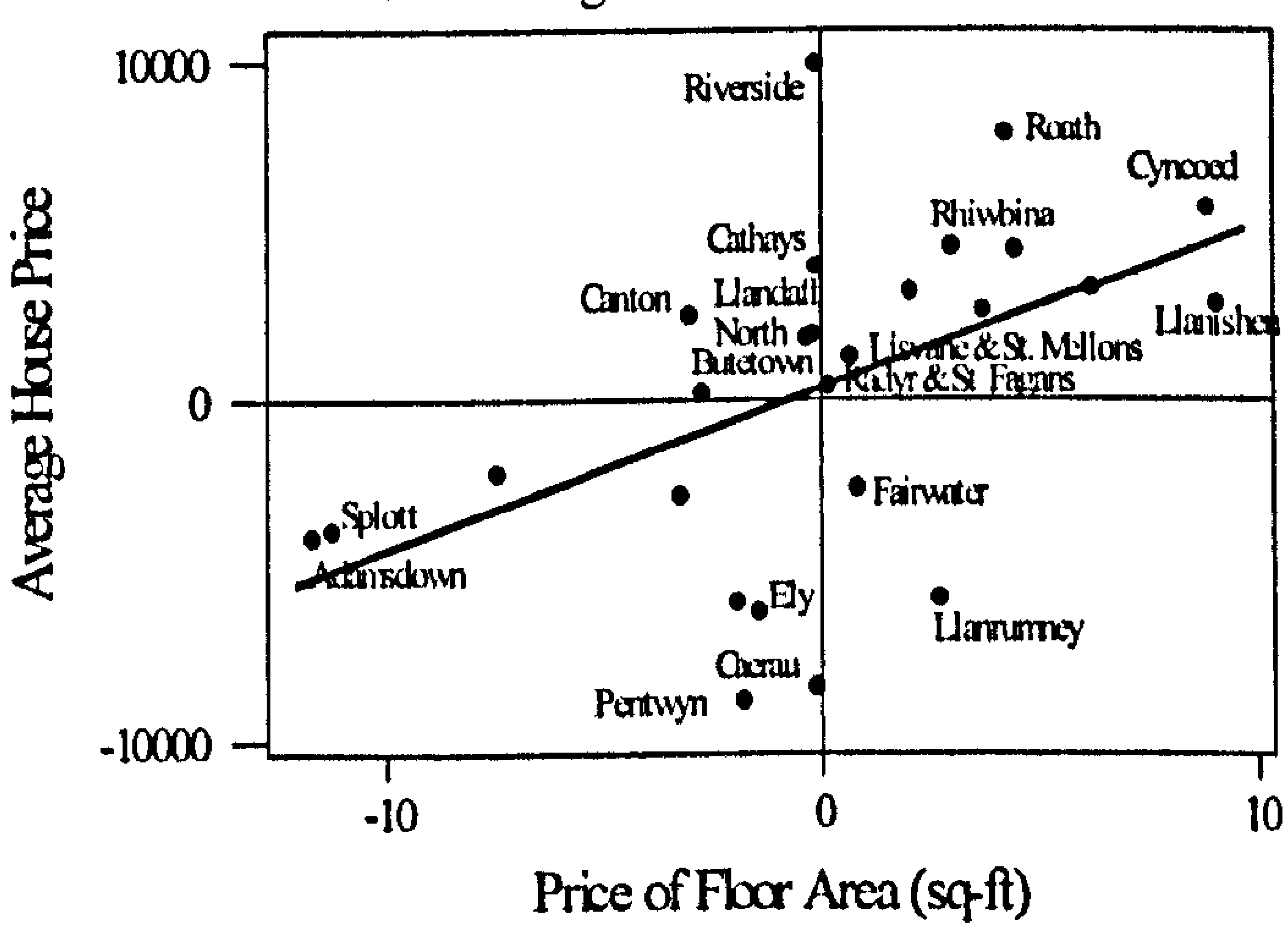


Figure 7.17
Relationship between Community Floor Area
and Average House Price Differentials

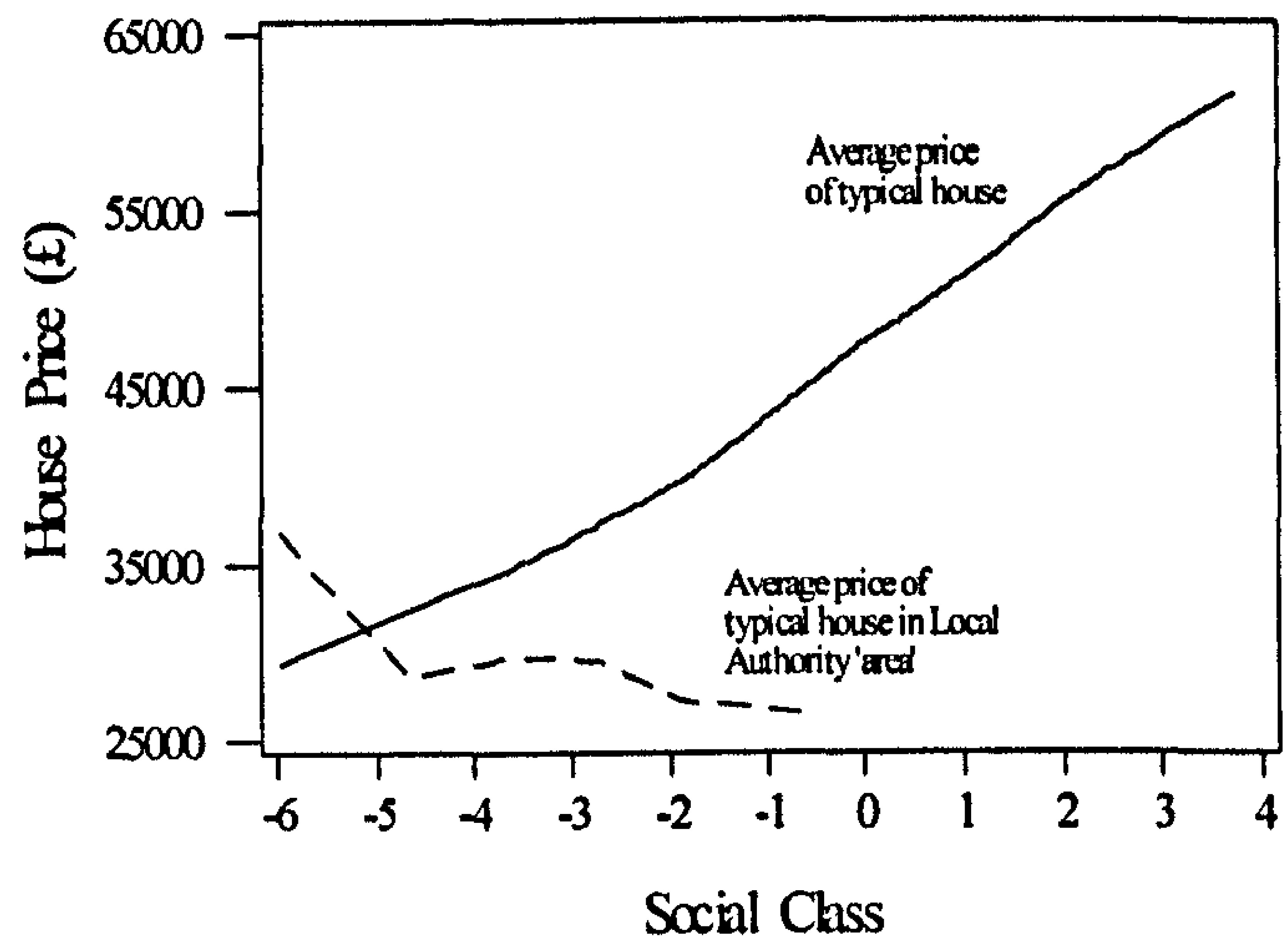


location externalities than buyers in Splott and Adamsdown. Hence, it can be expected that residents in the first three communities will be more anxious about the effects of negative externalities upon their property prices than the latter. A more interesting, and less trivial situation occurs in communities that do not fit into this general relationship. These tend to be located in the top, left-hand section of the graph (Riverside, Cathays, Canton, Llandaff North and Butetown) and the bottom right-hand side (Fairwater and Llanrumney). In the case of the former groups of communities, marginally more is spent upon location than upon the structural attributes of the property whilst the opposite is true for the latter two communities. In all these cases, any change in the effect of locational externalities will have a marginally bigger effect upon property prices than in the other communities. For instance, in the case of Riverside, Cathays and Canton, the proximity of Bute Park would appear to be having a beneficial effect upon property prices. Hence, any changes to Bute Park may have a negative effect upon prices in these communities. A similar argument can be made for Llandaff North (Llandaff Cathedral) and Butetown (Docklands). Conversely, property prices in the communities of Fairwater and Llanrumney are being depressed by negative locational externalities, and the improvement of these may increase property prices disproportionately compared to other communities. Finally, it is worth noting that the communities of Lisvane & St. Mellons and Radyr & St. Fagans correspond to the Cardiff average with respect to their structural and locational attribute prices. Hence, buyers in these communities are getting value for money, even though average house prices in these communities are the highest for the whole city.

Figure 7.18 describes the value of housing in Local Authority areas once community context has been taken into account. Once again, houses located in areas in which the majority of properties are Local Authority owned are more expensive *ceteris paribus*, compared to other areas of low social class. However, this premium has declined in absolute value, whilst the 'stigma effect' now occurs in areas of much lower socio-economic class (in this case, in areas with scores less than -5). This suggests that the spatial parameter drift models were under-estimating this 'stigma effect'. An examination of Figure 7.17 shows that the communities in which the majority of Local Authority areas are located (Ely, Caerau, and Pentwyn) have cheaper premiums than expected given the price of floor area, and hence it is the failure of the spatial drift specification to take into account these community effects that leads to this under-estimation. Figure 7.18 also shows how the multi-level specification allows the influence of social class to vary in a non-linear fashion, with the 'stigma-effect' initially decreasing with social class, before becoming influential again.

Figure 7.18

Local Authority Housing Stock and Social Class: Model 7.12



7.4.7 Conclusion

The multi-level specification would appear to best describe the spatial dynamics of the Cardiff housing market. By allowing house prices to vary across three levels, the interactions between the structural and locational attributes can be more precisely modelled. However, before any firm conclusion can be reached, tests for heteroscedasticty and spatial autocorrelation were undertaken to evaluate the degree to which the final model, Model 7.12, captures the spatial structures in the data. Table 7.17 is a summary of Breusch Pagan test statistics for each variable in the fixed part of the model. These suggest that heteroscedasticity is still present in the floor area and garden size attributes, although the effects are markedly reduced relative to those of Model 7.6. This heteroscedasticity is attributable either to variance heteroscedasticity in the attribute variables concerned, or the continued presence of spatial heteroscedasticity. The former would occur if higher income buyers spend marginally less on structural attributes and more on locational attributes than low income buyers. Hence, error variance associated with high income buyers would be different than their low income counterparts. Spatial heteroscedasticity is likely to occur since it is unlikely that the communities are correspond perfectly with submarkets, and hence communities will not be totally homogeneous with respect to supply and demand schedules. This will lead to structural instability of parameter estimates.

Figure 7.19

Standard Residuals of Model 7.12

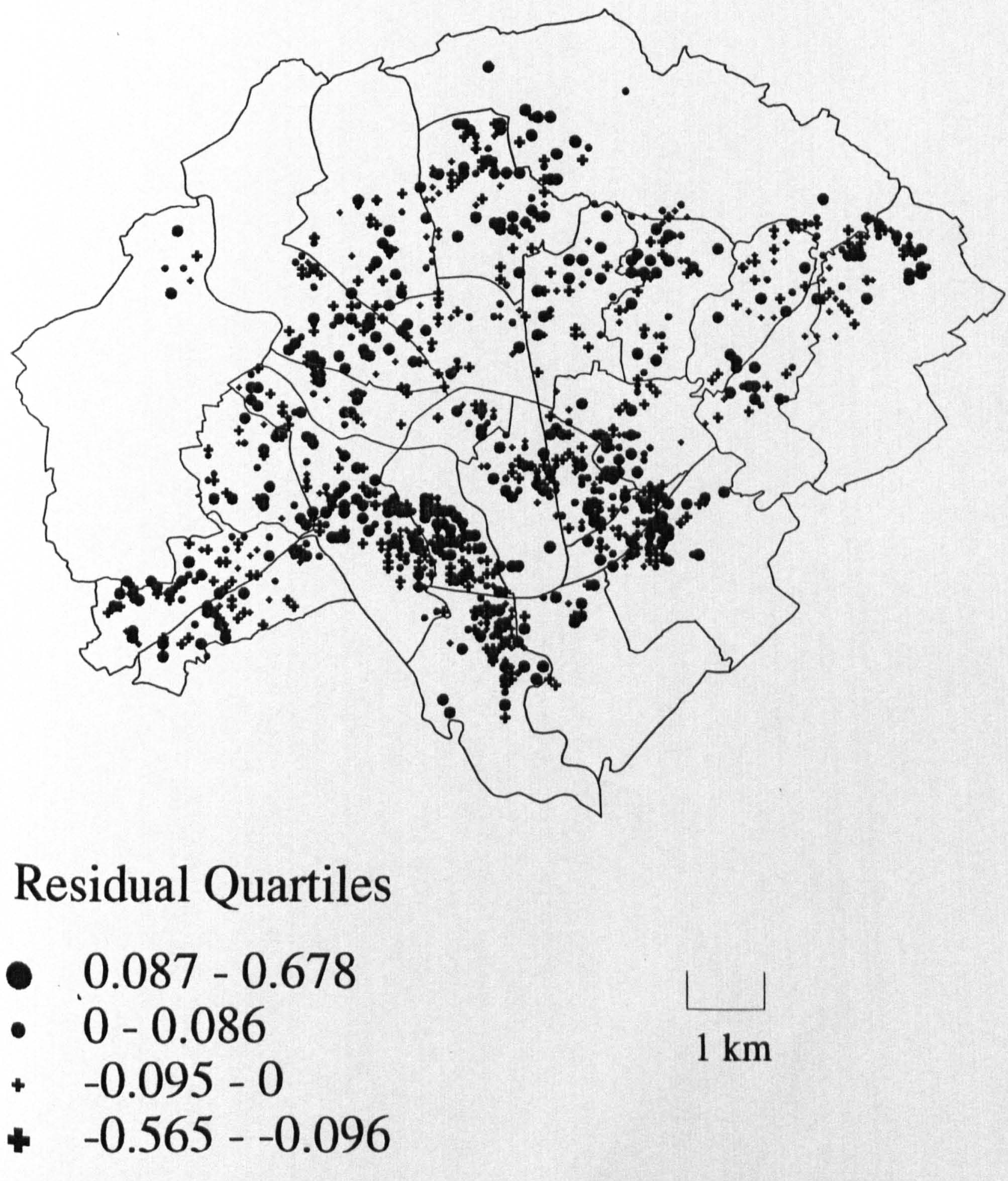


Table 7.17

Breusch-Pagan results for Model 7.12

Variable	Breusch-Pagan
Floor Area	21.36
Floor D	5.2
D Baths 2	0.003
D Shower 1	0.72
Full CH	0.58
Garage	4.6
ORP	1.02
Gdn:None	12.7
Gdn:5-50m	12.9
Gdn:>50m	0.9
Needs Mod	0.16
Dist CBD	0.87
Social	0.09
LA > 50%	3.7
SocLA	3.9

With regards to spatial autocorrelation, Figure 7.19 indicates that the spatial patterning of residuals from Model 7.12 are random, whilst the Moran I test confirms that spatial autocorrelation is no longer significant ($I = 0.00067$). This is to be expected however, since the multi-level specification will take into account spatial autocorrelation regardless of whether submarkets are correctly delimited. Hence, although the specification still needs slight improvements, it would appear that it has successfully captured the spatial structures in the data and therefore the spatial dynamics of the Cardiff housing market.

7.4.8 A Comparison of Specifications

It is clear that each specification conceptualises the Cardiff housing market as operating in a slightly different way. The ability of each to model the spatial dynamics of the housing market has relied upon theoretical considerations of how the housing market should operate, and diagnostic tests, to check how each model accounts for the spatial elements of the data. Taken together, these considerations suggest that the multi-level specification is the most suitable. However, this does not necessarily mean that the other specifications incorrectly estimate implicit prices. Rather, it may be the case that some housing attributes are robust to changes in specification. To ascertain the robustness of the models, the parameters of three of the models were compared - Model 7.3, Model 7.8 Model 7.16. These are the most sophisticated models for each of the specifications. The first thing to note is that, with the exception of floor area, Model 7.3 and Model 7.8 produce the most comparable parameter estimates. The estimates of Model 7.16 tend to be smaller. It is interesting to note the differences in the detached house variables. These are much smaller in the multi-level model, implying that these variables had captured locational differences in the previous models. Only the multi-level specification was able to model compositional and contextual effects simultaneously. There are several variables, such as garden size, which appear to be robust, and do not vary to a significant degree between the models. These tend to be structural attributes external to the physical dwelling, and this suggests that they may be valued differently to internal structural attributes.

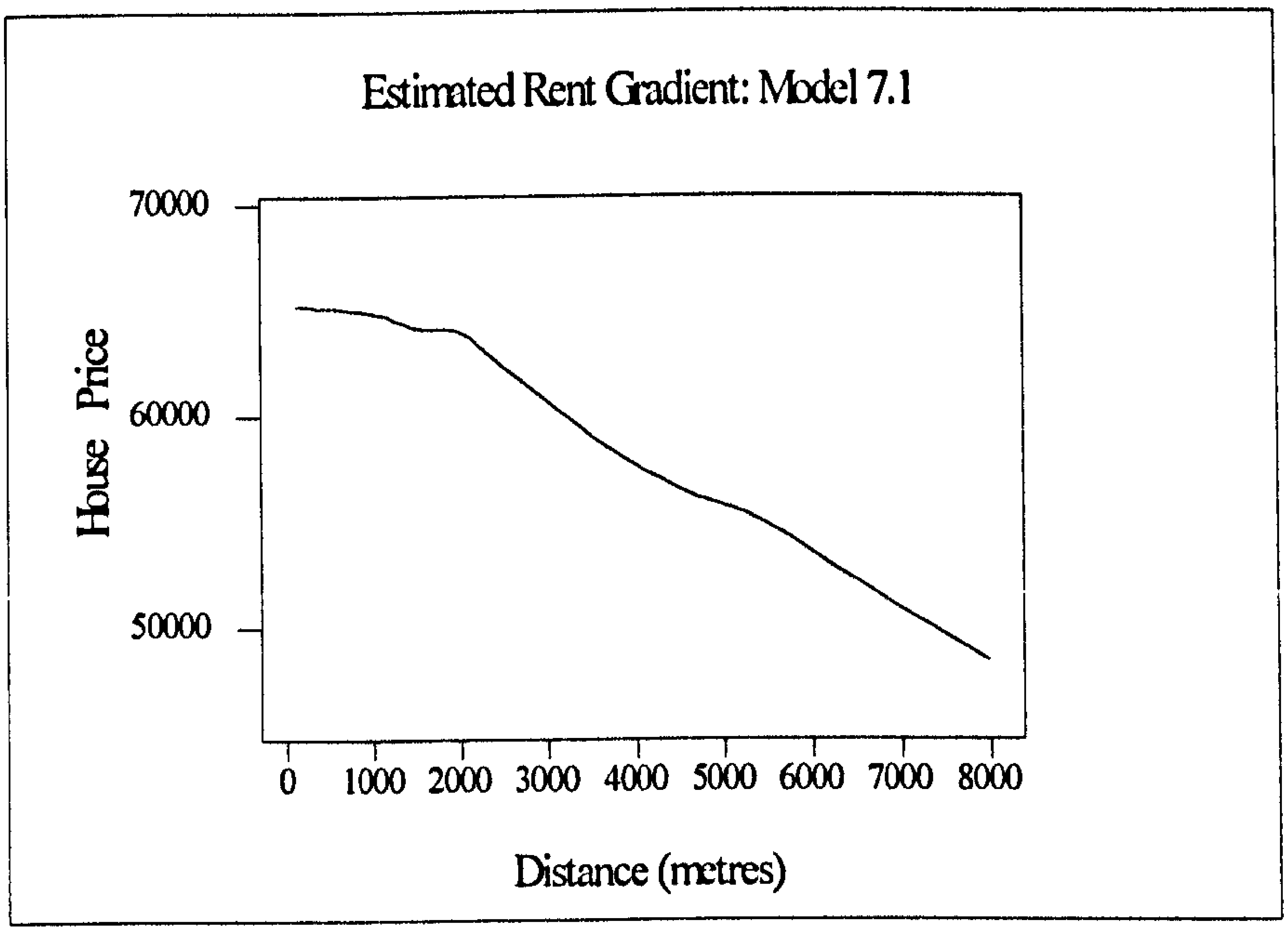
Section 7.5 The Cardiff Rent Gradient

One of the aims of this chapter was to estimate the rent gradient for Cardiff. The rent gradient is one of the basic tenets of standard urban economic theory, and hence its estimation is an important result, particularly in light of past research. Following the conclusions of *Chapter Six*, a linear functional form was used to estimate the rent gradient in all of the previous models. The resulting gradient gradients ranged from £0.70 to £1.90 per metre from the city centre, suggesting an average decline in property prices of around £130 per kilometre. Although this is quite a low value, Cardiff is a comparatively small city, and has relatively good transport routes from the city centre to the suburbs. This may result in accessibility not being valued particularly highly compared to other cities.

But as was argued in *Chapter Three,* a linear rent gradient is counter-intuitive, and a non-linear gradient, such as inverse power and negative exponential may be more appropriate, as postulated by urban economic theory. The problem lies in the fact that, despite what previous researchers have written, the exact functional form of the rent gradient cannot be known *a priori* . Also, there is no reason why the rent gradient should approximate any standard functional form. It was previously discussed in *Chapter Three* that many unique features of a city, such as transport routes, could distort the rent gradient such that it no longer fits a standard linear or non-linear form. Thus, to avoid constraining the model unnecessarily, five of the models were re-estimated using the dummy distance interval measures that were described in *Chapter Six*. The estimated implicit prices all had the

Figure 7.20

Estimated Rent Gradients for Selected Models



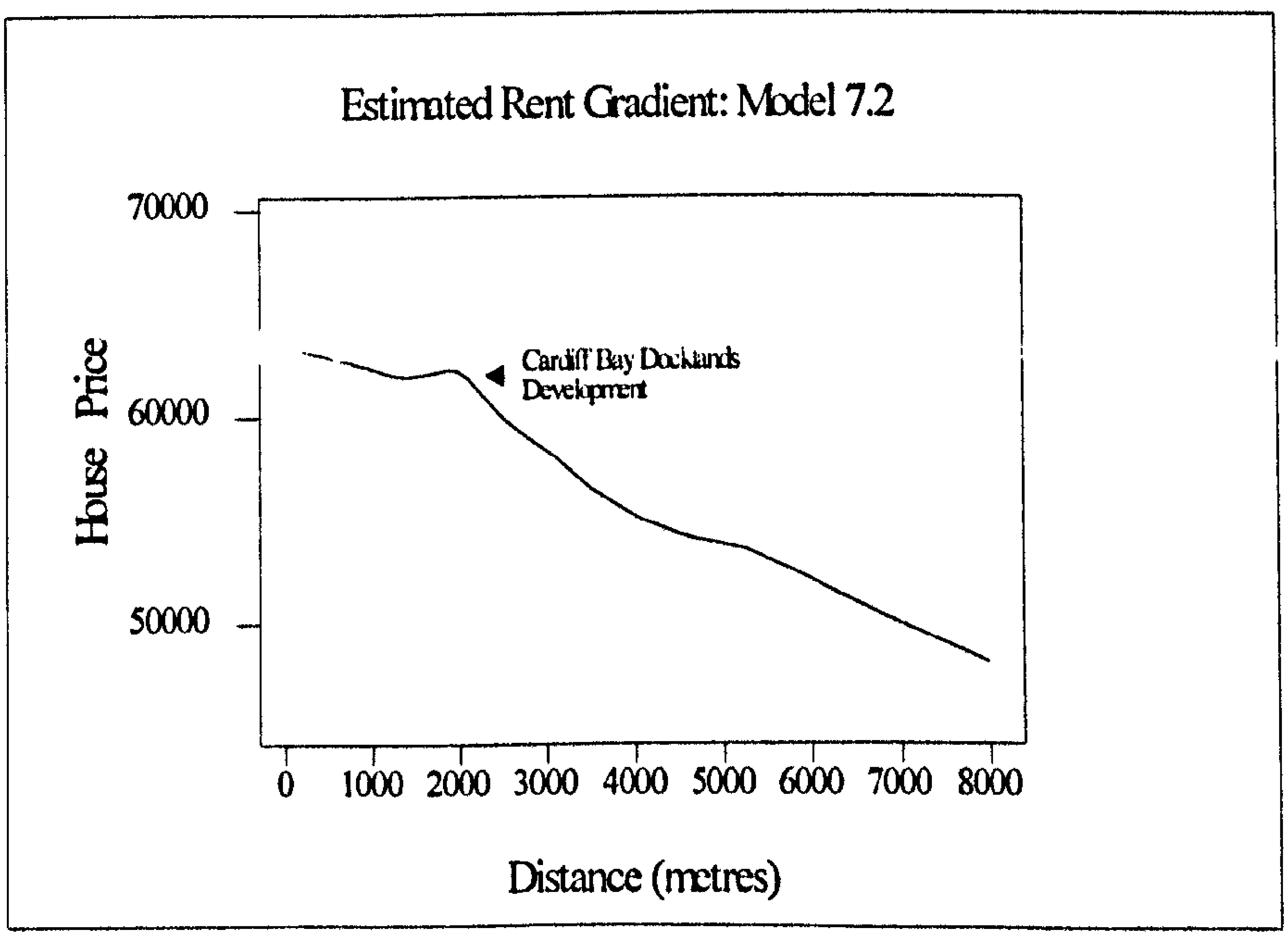
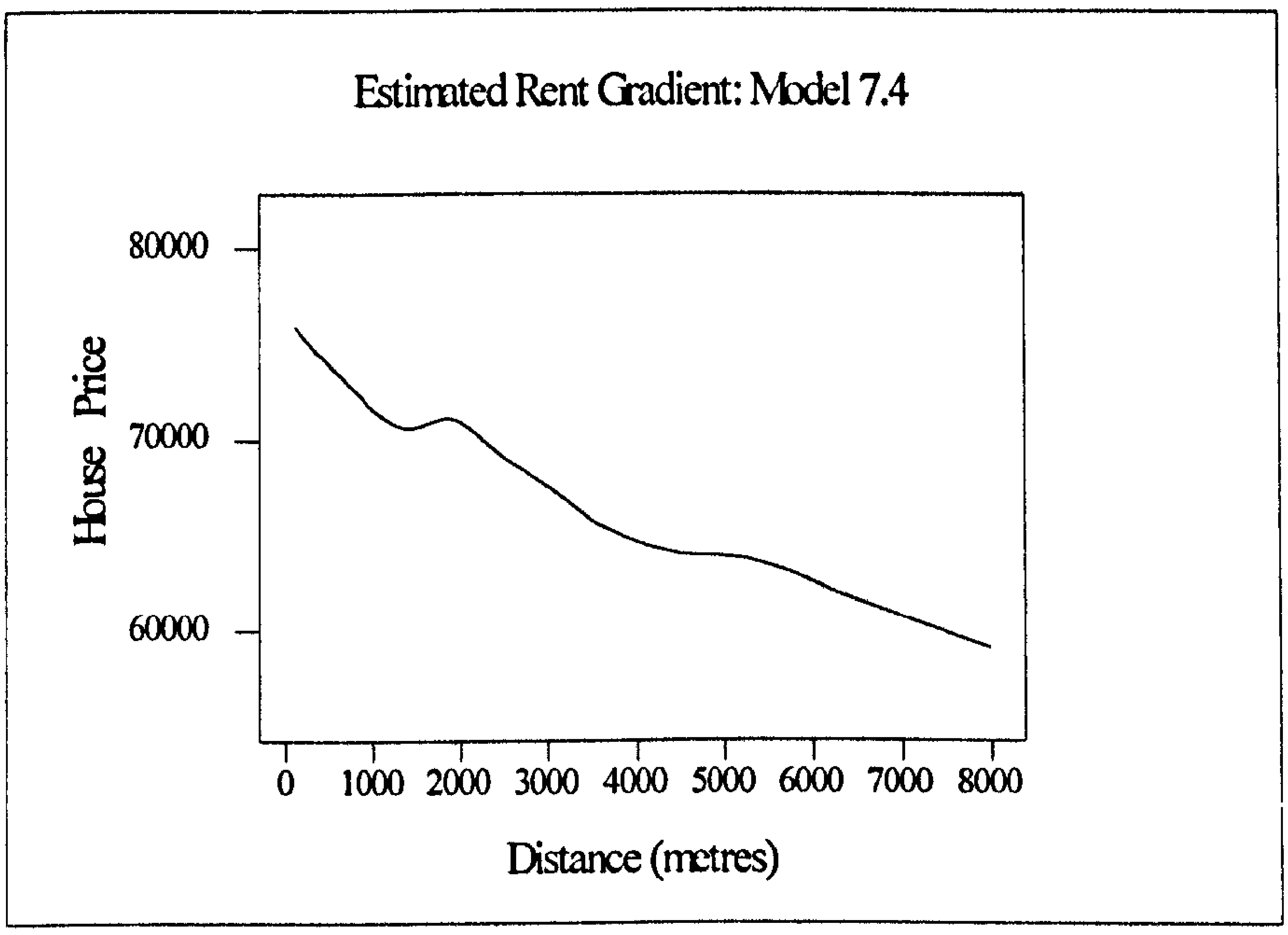
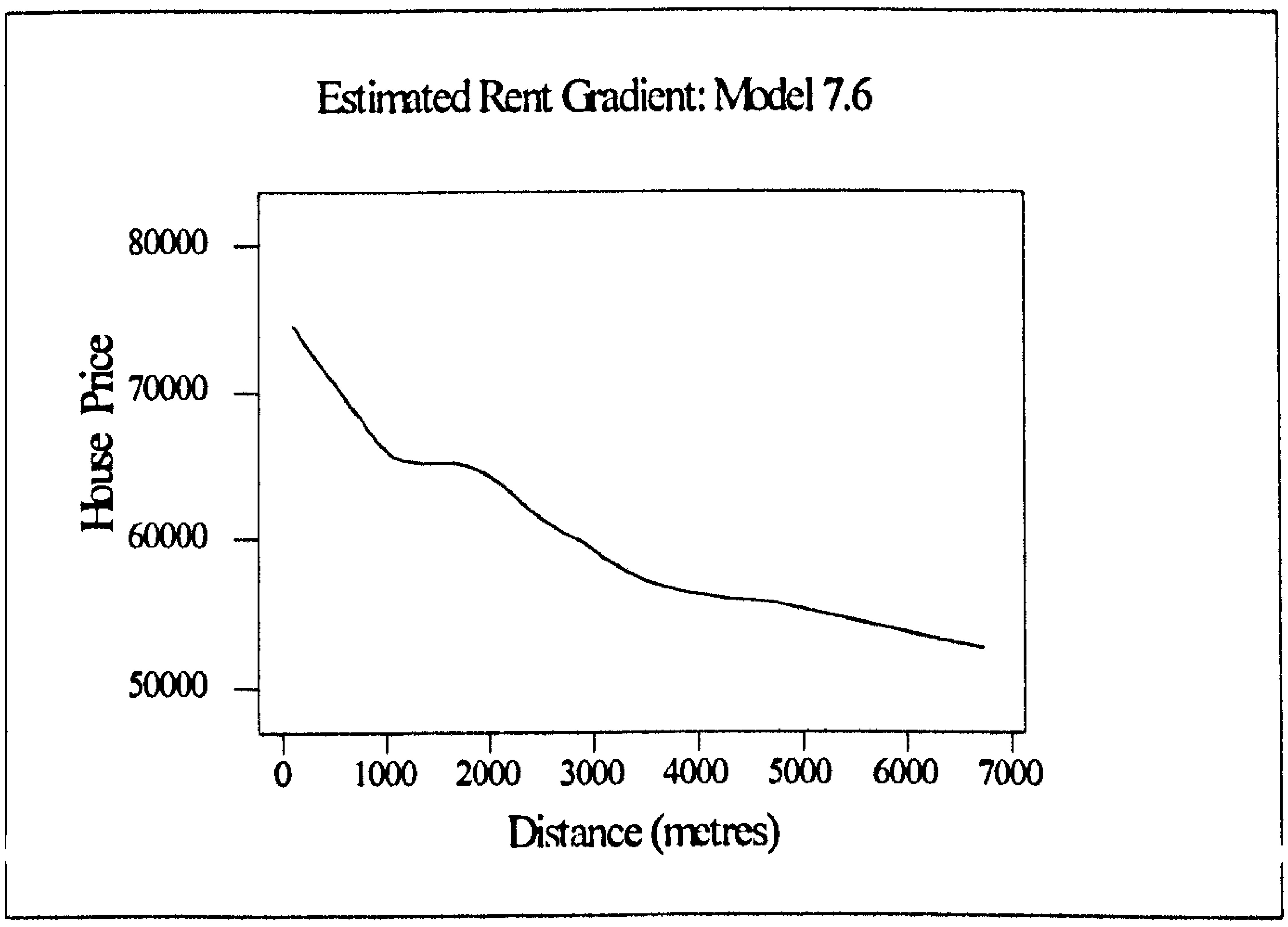
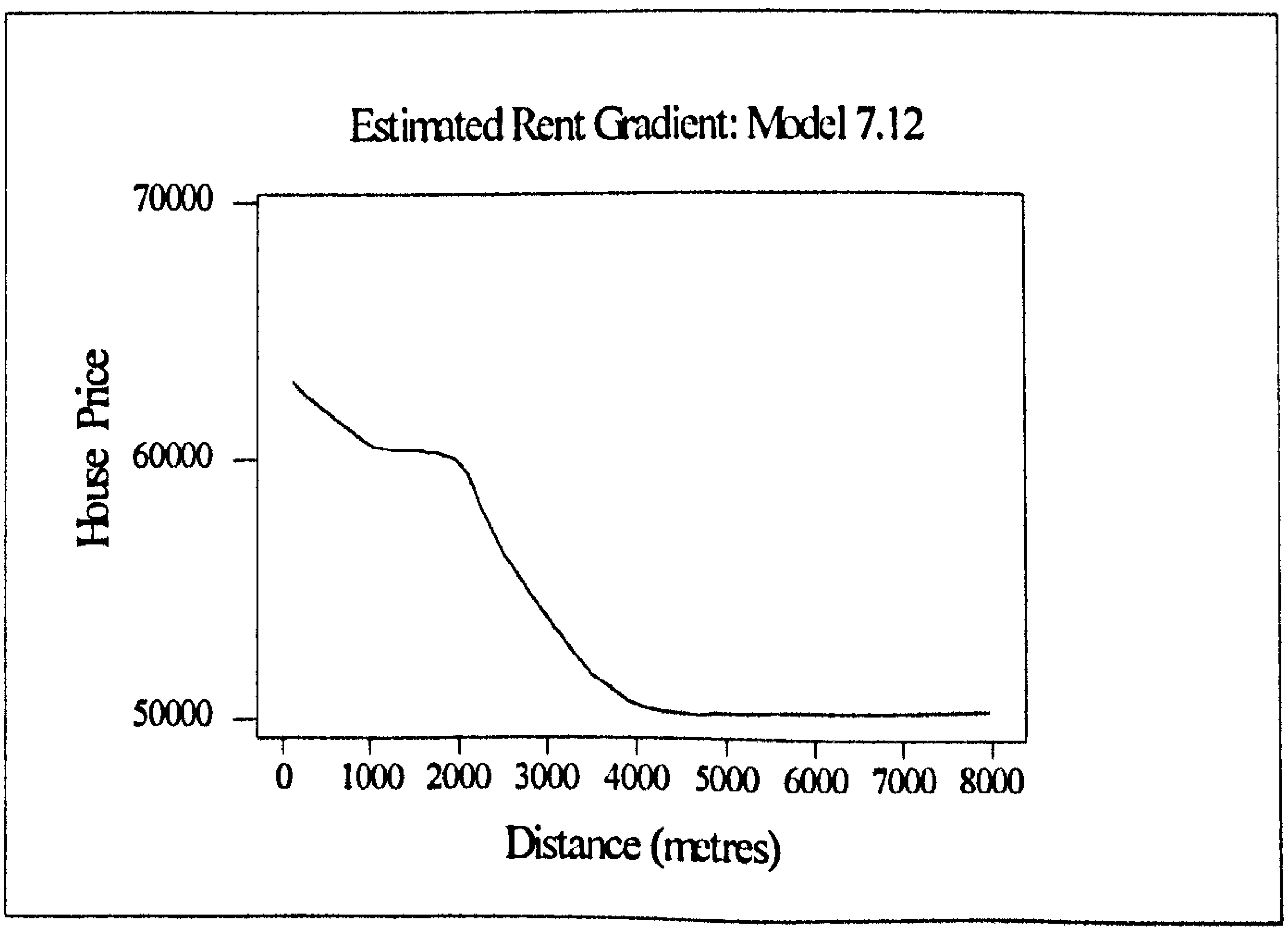


Figure 7.20 (cont.)







correct sign and were statistically significant. These were subtracted from their corresponding constant terms and plotted against distance using the locally weighted scatterplot smoother function to reveal the average price of the typical property at increasing distances from the city centre (Figure 7.20). Two important features are evident. Firstly, the rent gradients become less linear and increasingly concave from Model 7.1 through to Model 7.16. This is a very significant outcome, since it suggests that as the spatial dynamics of the housing market are more accurately modelled, the results conform more with urban economic theory. This result, coupled with the diagnostic tests, verifies that the multi-level specification best describes the structures of the Cardiff housing market. Secondly, all five graphs reveal a local maxima at around 2.5 km from the city centre. This coincides with the Cardiff Bay development area and this would imply that the rent gradient at this point is compensating for unspecified locational externalities associated with this area.

Section 7.6 Conclusions

The general aim of this chapter has been to model the spatial dynamics of the Cardiff Housing market. In doing so, several substantive results have been established. Firstly, on the technical and theoretical side, the three different specifications have been examined, and their ability to model spatial data evaluated. It has been shown that the ability to capture spatial effects has substantial influence upon the results. In particular, failure to take into account how structural attributes vary with locational context can lead to heteroscedasticity and spatial autocorrelation in the residuals, and incorrect implicit prices, such as the negative result for bungalows with two bathrooms. Moreover, the chapter has shown the advantages of modelling space using a multi-level specification as opposed to a single level one, such as the spatial parameter drift specification. In particular, the latter has shown the problems of modelling compositional and contextual effects, with the under-estimation of the 'stigma effect' in communities of predominately Local Authority tenure. The multi-level specification allowed both compositional and contextual effects to be modelled simultaneously, which re-evaluating the stigma-effect and allowing location attributes, such as social class, to vary with context. Spatial effects were still present in the final model, although this was a result of poor specification of submarkets.

The results also demonstrated the existence of submarkets, delimited according to both the housing stock and geographical area. By using the multi-level specification, the price of floor area - the most important structural attribute - was seen to vary significantly between communities, and housing type. The final model concluded that the price of floor area depended upon both the average price of a community and the social class of the area, and that separate market conditions exist for detached houses, irrespective of location.

The results also confirmed the importance of locational attributes in determining house prices. Social class was consistently the second most important variable in the model, and was highly significant. When the structural attributes were allowed to vary with social class, it was demonstrated that the price of floor area in areas of high social class were marginally less than in average social class, implying that marginally more was spent upon locational attributes. The influence of location was also established in the implicit prices of communities, which were the greatest in areas that had a high degree of positive locational externalities, such as Roath, Cyncoed, Riverside and Llandaff North. This was corroborated with the results of the multi-level models, that highlighted several communities in which the unit price of floor area was lower than predicted given the average community house price. An examination of the communities implied that unaccounted for locational externalities were responsible. Finally, the rent gradient form the city centre was estimated using dummy distance measures. This revealed that the rent gradient became more concave as the Models captured the spatial effects of the data. The results also suggested that, after four kilometres form the city centre, the rent gradient becomes negligible.

To summarise, the chapter has concluded that the multi-level specification best captures the spatial structures of the housing market, and has highlighted the importance of submarkets and locational attributes in determining house prices. This shall now be examined in more detail in *Chapter Eight*, in which the influence of specific locational attributes shall be modelled for properties in the Inner Area of Cardiff.

Chapter Eight

Towards a Valuation of Locational Externalities

Section 8.1 Introduction

The previous chapter has explored the spatial dynamics of the Cardiff housing market, from which three substantive conclusions were drawn. Firstly, it was shown that a multi-level specification best captured the complex geography of house price variations. In particular, the specification allowed the contextual effects of location and the compositional nature of the housing stock to be simultaneously considered. Secondly, the corollary of this result was the identification of submarkets operating across the housing market, delineated by both the housing stock and geographical area. The third conclusion concerned the importance of locational externalities as influential factors operating in the housing market. The results revealed that social class, a surrogate for locational externalities, was the most influential factor after house size. Moreover, it was demonstrated that in areas of 'very high' social class, locational externalities became marginally more important than floor area in determining property prices.

This chapter aims to expand on the last of these findings by evaluating the influence of specific locational externalities. As was discussed in *Chapters Five & Six*, the Inner Area of Cardiff was chosen as a study area to undertake this detailed spatial analysis. Building upon the results of *Chapter Seven*, the multi-level specification was used, and four levels of resolution were defined: the property level, the sub-street level, the HCS area level and the community level. These are described in more detail in *Chapter Five*, which also describes the construction of the context-sensitive GIS used to generate the locational attributes. The results of *Chapter Seven* also informed the manner in which the structural attributes entered the specification. To recap, Table 8.1 is a summary of the locational attributes that are hypothesized to influence house prices in the Inner Area. These have been grouped according to the spatial level at which they are conceptualised to operate. In comparison to *Chapter Seven*, social class has now been operationalised at the community (ward) level, since it is no longer acting as a surrogate variable for locational attributes, but as a

Table 8.1
Inner Area Locational Attributes

Variable	Description	Abbreviations
Property Level		
Accessibility to CBD	Continuous	Dist CBD
Accessibility to M4 motorway	Continuous	Dist Mway
Accessibility to railway stations	Continuous	Dist Station
Proximity to hospitals	Continuous	Hospital
Proximity to sports centres	Continuous	Sports
Proximity to community centres	Continuous	Community
Proximity to institutional centres	Continuous	Institutional
Proximity to local shops	Continuous	Shops
Proximity to primary schools	Continuous	Primary
Proximity to secondary schools	Continuous	Secondary
Proximity to Bute Park	Continuous	Bute Park
Proximity to parks / open space	Continuous	Parks
Proximity to light industrial land-use	Continuous	Light Ind
Proximity to heavy industrial land-use	Continuous	Heavy Ind
Rail 0 -50m	Dummy	Rail 0-50m
Rail 50 - 100m	Dummy	Rail 50-100m
Rail 100 - 150m	Dummy	Rail 100-150m
Rail 150 - 200m	Dummy	Rail 150-200m
River 0 - 50m	Dummy	River 0-50m
River 50 - 100m	Dummy	River 50-100m
River 100 - 150m	Dummy	River 100-150m
River 150 - 200m	Dummy	River 150-200m
Sub-Street Level		
Road Type: Primary	Dummy	Primary Road
Road Type: Secondary	Dummy	Secondary Road
Road Type: Residential	Dummy	Residential Road
Road Type: Cul-de-sac / Close	Dummy	Close
Street quality 0-50m: Poor	Dummy	Poor 0-50m
Street quality 0-50m: Below Average	Dummy	Below Ave 0-50m
Street quality 0-50m: Above Average	Dummy	Above Ave 0-50m
Street quality 0-50m: Good	Dummy	Good 0-50m
Street quality 50-100m: Poor	Dummy	Poor 50-100m
Street quality 50-100m: Below Average	Dummy	Below Ave 50-100m
Street quality 50-100m: Above Average	Dummy	Above Ave 50-100m
Street quality 50-100m: Good	Dummy	Good 50-100m
Street quality 100-200m: Poor	Dummy	Poor 100-200m
Street quality 100-200m: Below Average	Dummy	Below Ave 100-200m
Street quality 100-200m: Above Average	Dummy	Above Ave 100-200m
Street quality 100-200m: Good	Dummy	Good 100-200m
Street non-residential buildings.	Dummy	Non-res Buildings
Sch Catchment: Willows High School	Dummy	Sch: Willows
Sch Catchment: Fitzalan High School	Dummy	Sch: Fitzalan
Sch Catchment: Cantonia High School	Dummy	Sch: Cantonia
Sch Catchment: Cathays High School	Dummy	Sch: Cathays
Sch Catchment: St Teilo's High School	Dummy	Sch: St Teilo's

Table 8.1 (cont.)

Variable	Description	Abbreviations
HCS Area Level		
Percentage Local Authority tenure	Dummy	La > 50%
Percentage of open space	Continuous	% Open Space
Percentage of non-residential land-use	Continuous	% Non-Residential
Housing density	Continuous	Density
Quality of local shops	Continuous	Q.Shop
Quality of local public transport	Continuous	Q.Transport
Quality of local sport facilities	Continuous	Q.Sport
Quality of local parks	Continuous	Q.Parks
Quality of local community facilities	Continuous	Q.Commuinty
Community Level		
Social class	Continuous	Social

contextual background against which supply and demand mechanisms operate. In contrast, the variable measuring areas of predominately Local Authority tenure were aggregated to the HCS Area Level, reflecting the very localised nature of this tenure within the Inner Area.

This chapter is divided into four sections. The second section builds up a simple random intercepts model, which assumes that the implicit prices are stable across the Inner Area and the housing market operates as a unified whole. Section three then allows the housing attributes to vary at the higher levels to simulate the spatial nature of the housing market. The final section concludes the chapter.

Section 8.2. The Spatially Uniform Housing Market.

This is the random intercepts model, that allows the average price of property (i.e. the constant terms) to vary at the different levels of resolution, but keeps the implicit prices of the housing attributes constant. This amounts to an assumption of a spatially homogeneous housing market, where supply and demand schedules are uniform. The estimation of the model is divided into six parts, each describing the effect of including specific housing attributes at the various levels.

8.2.1 Model 8.1 The Grand Mean Model

Table 8.2 summarises this basic model, which estimates a grand mean house price for the whole of the Inner Area of around £53 500. The variation around this average price has been decomposed into variation at the level of the property, the street, the HCS area and the community. With respect to each level, the greatest proportion of the total variation in house price occurs between communities (£6828), with the least between HCS areas (£2033) within a given community. The likelihood ratio statistic was calculated so the effects of additional terms in the model can be judged.

Table 8.2

Model 8.1 - The Grand Means Model

FIXED

PARAMETER	Coeffient	S.Error	T-stat
CONSTANT	53637	423.68	126.6

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	6829	3129.83	2.18
HCS Area Level			
CONSTANT	2033	637.61	3.19
Sub-Street Level			
CONSTANT	3640	672.61	5.41
Property Level			
CONSTANT	4903	493.07	9.94

-2*(log-likelihood) = 317.04

Table 8.3

Model 8.2 - The Structural Attributes Model

FIXED

PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	43477	218.74	198.76
FLOOR AREA	31.18	1.43	21.80
FLOOR D	8.66	2.97	2.92
MT BATH 2	9807	2987.88	3.28
FULL CH	3751	830.39	4.52
GARAGE	3933	883.78	4.45
ORP	4173	1008.99	4.14
GDN:NONE	-4334	1111.06	-3.90
GDN:5-50M	2657	972.46	2.73
NEEDS MODS	-3698	1091.62	-3.39

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	1461	723.20	2.02
HCS Area Level			
CONSTANT	910	238.63	3.81
Sub-Street Level			
CONSTANT	830	196.81	4.22
Property Level			
CONSTANT	1620	163.79	9.89

-2*(log-likelihood) = -339.123

Key

Abbreviation	Variable
FLOOR AREA	Total Floor Area (sq-ft)
FLOOR D	Total Floor Area Detached Housing
MT BATH 2	Mid-Terrace Two Bathrooms
FULL CH	Full Central Heating
GARAGE	Number of Garages
ORP	Off-Road Parking
GDN:NONE	Garden:None
GDN:5-50M	Garden:5-50 metres
NEEDS MODS	In Need of Modernisation

8.2.2 Model 8.2 The Structural Attributes Model

To isolate the effects of locational externalities, the structural attributes were first added to the grand means model to account for the compositional effects of the housing stock (Table 8.3). Due to the similarity of the housing stock, the number of significant variables in the model are reduced compared to the models in *Chapter Seven*. For instance, there are very few properties with gardens greater than fifty metres in length. The constant term now represents the price of an averaged sized terraced property (£43477). The most influential variable is again floor area, with the model also depicting the separate market conditions for detached housing. The variable measuring shower rooms was insignificant, whilst the value of modernising a property was estimated to be around £3700.

The inclusion of the structural attributes has resulted in a large decline in the variance at all levels, but particularly at the community level, suggesting that the differences between communities in Model 8.1 was caused principally by differences in the housing stock. This has resulted in the property level and the community level explaining roughly the same amount of variation in house price. Under the null hypothesis, the difference of the likelihood's (656.16) follows a chi-square distribution with 9 degrees of freedom. The probability of obtaining a chi-square of this magnitude by chance is negligible (less than 0.001), strongly indicating that the structural attributes have an important effect in explaining house price variation in the model.

8.2.3 Model 8.3 The Property Level Locational Attributes Model

The previous model has taken into account the compositional effects of the housing stock. It is now time to evaluate the effect of the locational attributes at each of the four levels, starting with the level of the individual property. The locational attributes that operate at this level are those associated with accessibility to work place and proximity effects to non-residential landuses.

8.2.3.1 Accessibility

Three measures of accessibility - access to the CBD, access to the nearest motorway junction and access to the nearest railway station - were modelled using measures generated by the *NETWORK* tool in ARC/INFO, as described in *Chapter Six*. In accordance to the findings of *Chapter Seven*, a linear functional form was used for accessibility to the CBD and the motorway, although a negative exponential was used for accessibility to the railway stations, since it is hypothesized that this will only be significant over very short distances. Table 8.4 is a summary of the estimated parameters. It can be seen that only accessibility to the nearest motorway junction is significant. The insignificance of the remaining two can be explained by the fact that the majority of properties are within close enough proximity to make accessibility to the CBD and railway stations relatively unimportant. However, it is interesting to note the similarity between the parameter estimate for accessibility to the motorway, and that for accessibility to the CBD estimated in *Chapter Seven*.

Table 8.4

Accessibility Parameter Estimates for Model 8.3

PARAMETER	Coefficient	S.Error	T-stat
DIST CBD	-1.52	1.53	-0.99
DIST MWAY	-2.65	1.21	-2.19
DIST STATION ($\beta = -20$)	4.06	3.32	1.22

8.2.3.2 Proximity to Non-residential Landuse

Chapter Six described how two types of proximity measures to various landuses were generated in ARC/INFO. The first employed the use of dummy variables to capture the effects of proximity at various distances from specific externalities, namely railway lines and rivers / water fronts . These were generated in ARC/INFO by the use of buffer zones. Table 8.5 is a summary of the parameter estimates of the model.

Table 8.5

Railway Lines and River Taff Parameter Estimates

PARAMETER	Coefficient	S.Error	T-stat
RAIL 0- 50M	-2790	1401.55	-2.02
RAIL 50-100M	-86	1343.09	-0.06
RAIL 50-150M	-10645	1288.78	-0.83
RAIL 150-200M	-793	1292.45	-0.61
RIVER 0-50M	8866	2747.14	3.23
RIVER 50-100M	-4912	2665.45	-1.84
RIVER 100-150M	-1233	1878.92	-0.66
RIVER 150-200M	-1891	1698.70	-1.11

It can be seen that, for both railway lines and rivers, only properties within fifty metres are significantly affected. Since these represent properties directly facing them, it can be regarded as an aesthetic cost. The model suggests that railway lines reduce the value of a property by around £2800, whilst close proximity to river will increase the value of a property by nearly £9000. This is a very high value and may be a surrogate for lower densities and / or access to parkland given the propensity of the River Taff to flood. As such, these values shall be re-evaluated in a later section.

The second measure of proximity was based upon continuous distance from the externality, taking into account the underlying topology. These were generated using the ACCESSIBILITY command. This command incorporates an attractiveness index to model the magnitude of the externality, and a distance decay function to model proximity (see equation 6.1) . For areas of non-residential landuse, such as industrial sites and parkland, the magnitude was calculated as the area of land squared. For other non-residential landuses, such as shops and schools, the attractiveness index was set to unity. Since the shape of the distance decay function was not known *a priori*, five β-values, ranging from 0.25 - 3.0, were used to estimate five distance decay functions. This is known as calibration, and is an important part of modelling externality effects Waddell, et al (1993). A small β-value represents a gentle distance decay curve and hence the greater the extent of the effect. A large β-value represent a steep distance decay curve, and the externality only has an effect over a short distance. The aim of the research is not to find the exact β-value for each externality *per se*, but to discover over what range of β-values the externality effect is significant, and then to compare the relative effects of each externality. Thus, each of the five externality measures was modelled separately, and the significance of the parameter used to determine the appropriateness the of the β-value.

Table 8.6 is a summary of the t-statistics of the estimated externalities. An insignificant result for all values suggests that the externality has a steep distance decay function and hence has a negligible effect upon property prices. Conversely, the larger the t-statistic, the better the β-value captures the effect of the externality. For instance, it can be seen that the effect of Bute Park is significant between the range of 0.25-2.0. However, the t-statistics decrease in magnitude as the β-values increase. This suggests that Bute Park has a gentle distance decay function, best captured by a small β-value.

The insignificance of the t-statistics for the overall effects of non-residential landuse confirms that property prices are not influenced by landuse *per se*, but by specific externalities. As was described in *Chapter Six*, for this reason non-residential landuses were separated into various classifications of landuses, the primary classifications being industrial

Table 8.6

T-Statistics for Non-residential Landuse Proximity Estimates

β	0.25	0.5	1.0	2.0	3.0
Non-residential Landuse	0.366	0.088	0.364	0.96	1.74
Bute Park	3.86	3.604	2.93	2.00	1.69
Parks	3.318	2.78	2.24	1.06	0.68
Industrial	0.02	0.05	0.05	1.114	0.8
Industrial: Heavy	2.01	2.27	1.51	0.48	0.11
Industrial: Light	0.295	0.467	0.277	1.03	0.763
Community	0.90	1.59	1.97	1.94	1.92
Institutional	0.22	0.668	0.23	0.131	0.01
Hospital	1.179	0.98	0.61	0.26	0.29
Sports	0.875	0.016	1.207	2.09	2.06
Shops	0.285	0.27	0.015	0.121	0.010
Primary School	1.83	1.76	1.77	2.03	2.14
Secondary School	1.66	1.56	1.30	0.878	0.692

sites and open space. The t-statistics for parks follow a similar pattern to Bute Park, although they are smaller in magnitude, suggesting less of an impact. The effect of industrial landuse has an insignificant impact upon property prices until the distinction is made between 'light' industrial areas and 'heavy' industrial areas. Heavy industrial areas have the most significant effect with a distance decay β-value of around 0.5, becoming insignificant at values greater or less than this. This implies that although the effect of heavy industrial areas is quite extensive, they are not as extensive as parks or open space. Hence, it may be the case that sellers of property accentuate proximity to positive externalities, such as parks and playdown proximity to negative externalities, such as heavy industrial areas. The t-statistics suggest that proximity to light industrial areas have a negligible effect on property prices, although this may be a result of the fact that the sample contained very few properties significantly near these areas.

The remaining externalities were estimated with an attractiveness index set to unity and hence the externality effect is solely determined by proximity. The effects of proximity to

Table 8.7

Model 8.3 - Property Level Locational Attributes

FIXED

PARAMETER	Coeffient	S.Error	T-stat
CONSTANT	43537	222.31	195.84
FLOOR AREA	31.01	1.39	22.32
FLOOR D	8.71	2.86	3.04
MT BATH 2	7524	2811.18	2.68
FULL CH	4000	797.75	5.01
GARAGE	3947	849.61	4.65
ORP	3662	963.10	3.80
GDN:NONE	-4885	1080.35	-4.52
GDN:5-50M	2255	924.37	2.44
NEEDS MODS	-3656	1051.65	-3.47
DIST MWAY	-2.65	1.21	-2.19
BUTE PARK($\beta = 0.25$)	11644	3016.68	3.86
PARKS ($\beta = 0.25$)	159256	47997.53	3.32
HEAVY IND($\beta = 0.5$)	-24960	10979.76	-2.27
SPORTS ($\beta = 2.0$)	0.0002236	0.0001070	2.09
PRIMARY($\beta = 3$)	0.0004350	0.0002030	2.14
RAIL 0-50M	-2790	1401.55	-2.02
RIVER 0-50M	8869	2747.14	3.23

RANDOM

PARAMTER	Coefficent	S.Error	T-stat
Community Level			
CONSTANT	944	469.54	2.01
HCS Area Level			
CONSTANT	468	153.68	3.05
Sub-Street Level			
CONSTANT	718	177.75	4.04
Property Level			
CONSTANT	1503	150.90	9.96

-2*(log-likelihood) = -406.374

community centres has a β-value of around 1.0, at which it is border-line significant. This implies that the externality effect is directly proportional to distance from it. Institutional centres are insignificant for the whole range of β-values, suggesting that they have very little impact of property prices. The externality effects of hospitals are also insignificant for the range of β-values. However, there is a trend of the t-statistics increasing with decreasing β-values, indicating that the externality effect may become significant for very small values of b. This hypothesis was tested using β-values within the range of 0.125 - 0.0156, which produced parameters with t-statistics of between 1.71 - 1.89, bordering upon significance. This implies that hospitals have a very gentle distance decay curve, influencing property prices across a much wider area than that being studied. If this is the case, then the Inner Area will be too small to capture this effect.

The results for proximity to sport centres implies that their externality effects operate across small distances, with its optimal β-value falling within the range of 2.0 - 3.0. This suggests that sport centres are only influential if they are within walking distance. A similar result applies to proximity to primary schools, but not secondary schools. This is an interesting result since it suggests that proximity to schools is only a consideration for households with children of primary school age. Of course, the catchment area within which a property falls may be important with respect to secondary schools, and this is evaluated in the section 8.2.4 Finally, the insignificant results for shops can be explained by the close proximity to the city centre and the high density of smaller shopping areas in the Inner Area.

8.2.3.3 The Externality Parameter Estimates

Table 8.7 presents Model 8.3 with the significant property level locational attributes. The addition of these attributes has had the greatest effect upon the bathroom variable, implying that this was capturing the effect of positive externalities, in this case proximity to Bute Park. The random terms in the model suggests that the addition of the property level locational attributes has had the greatest effect of explaining higher level variation, particularly at the HCS area and community level, with only a negligible effect upon the variation between properties. This is understandable since at the property level, the majority of variation is caused by differences in structural attributes, whilst locational attributes would tend to be quite similar. The chi-squared test indicates that the difference of the likelihood statistics (67.25) is significant at the 99% level with 8 degrees of freedom,

Figure 8.1

Valuing Proximity to Heavy Industrial Areas: Model 8.3

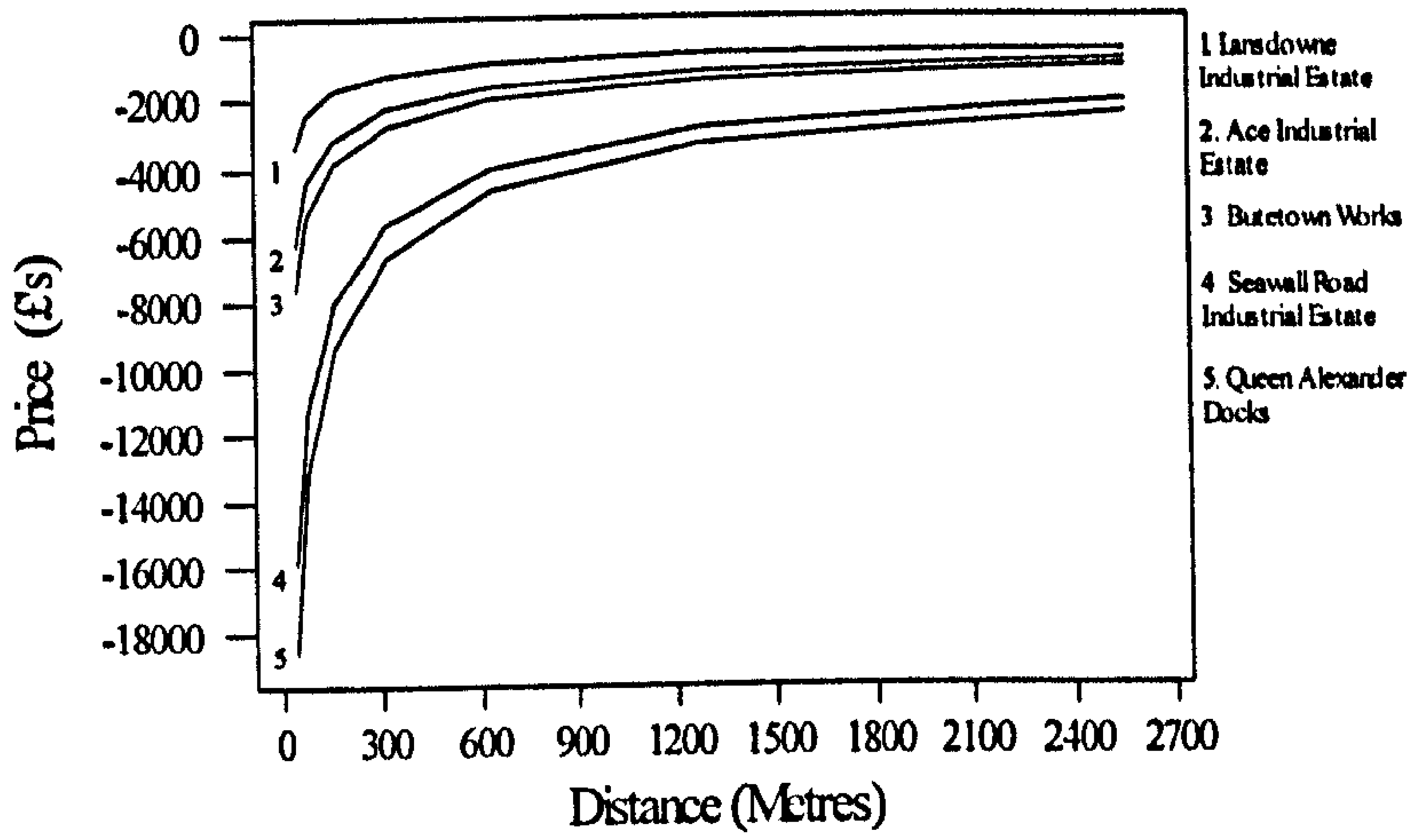


Figure 8.2

Valuing Proximity to Parks and Open Space: Model 8.3

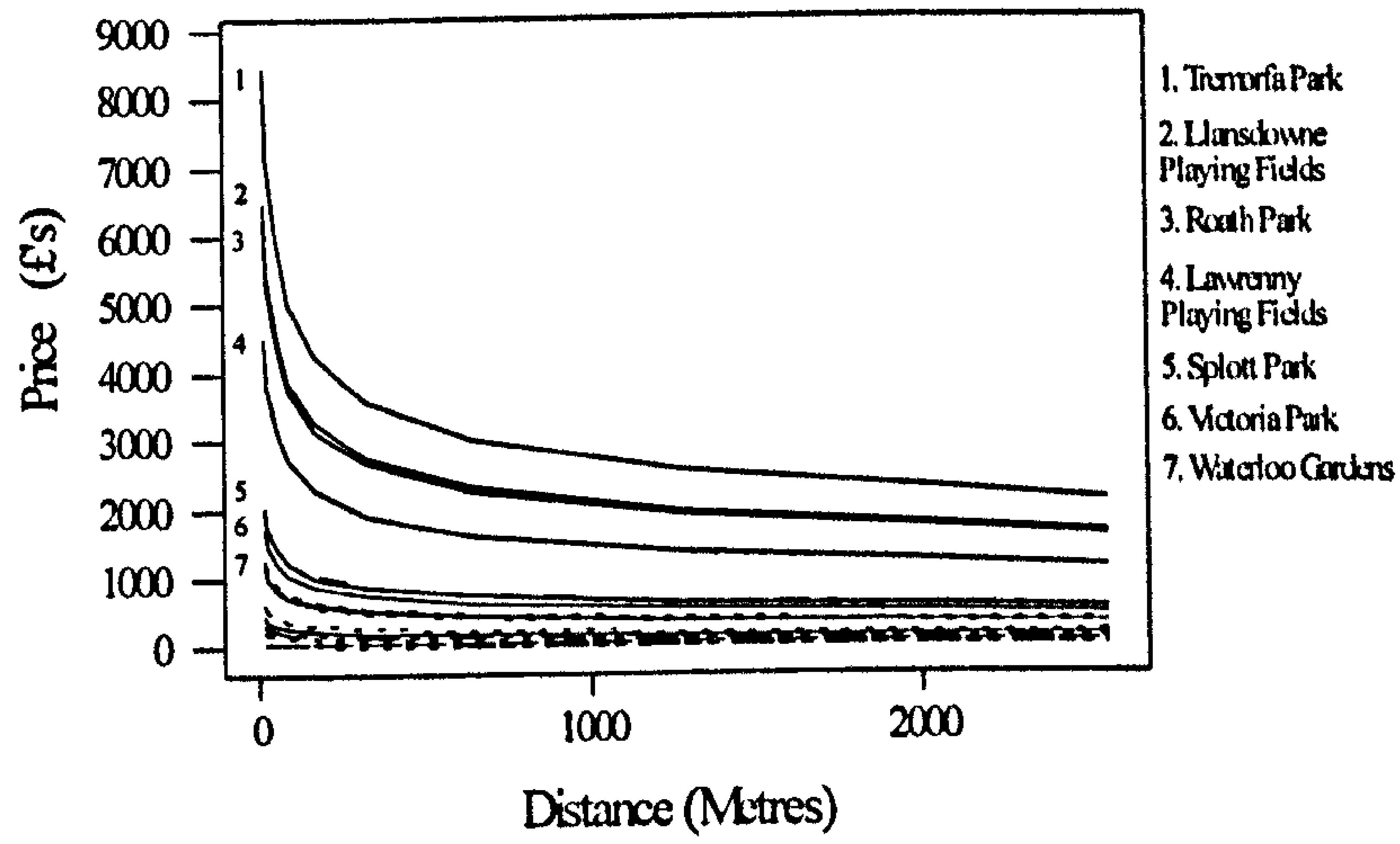
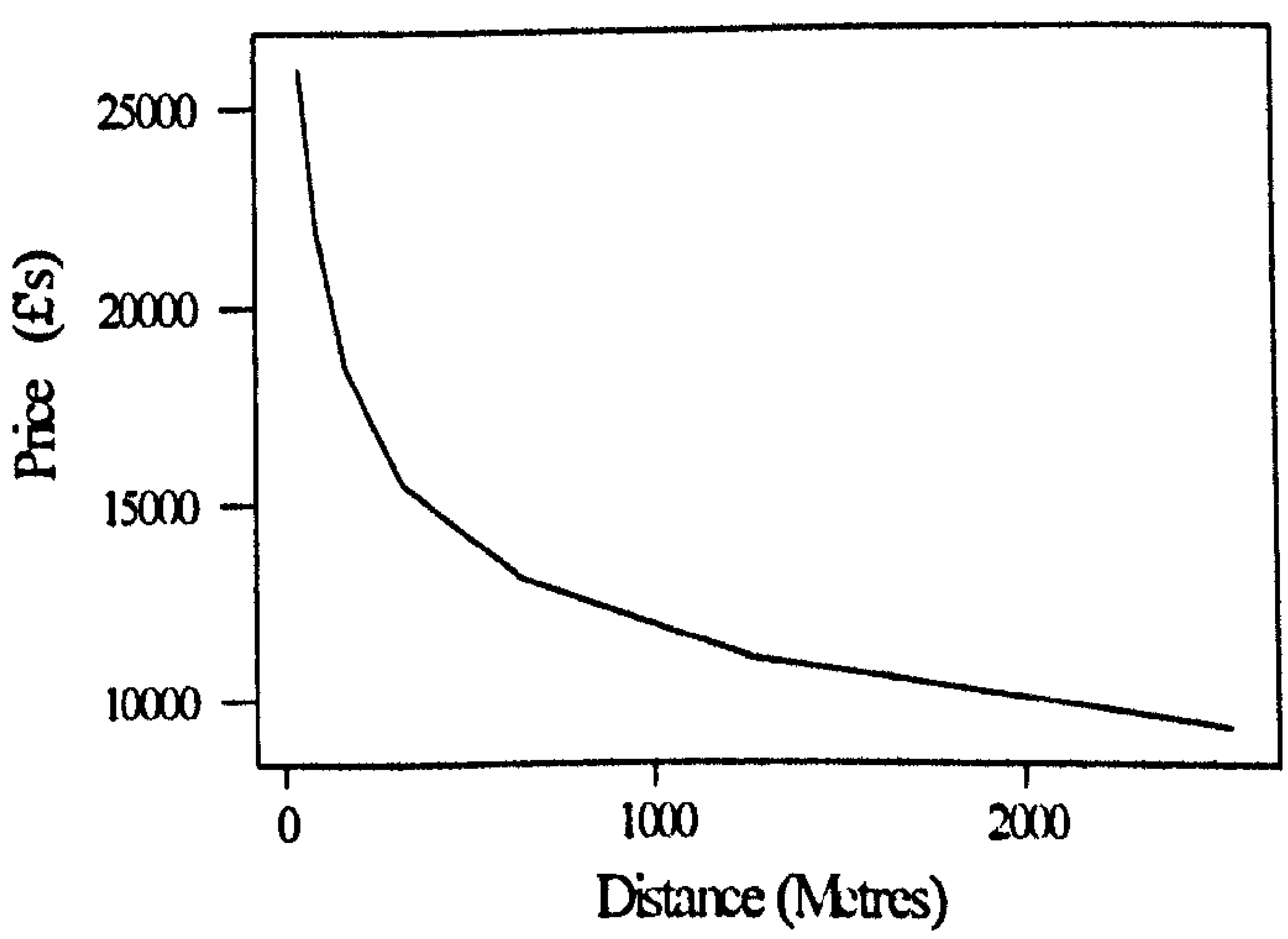


Figure 8.3

Valuing Proximity to Bute Park: Model 8.3



suggesting that the property level locational attributes have an important statistical effect in explaining house price variation.

8.2.3.4 The Magnitude and Geography of Property Level Externalities

Looking at the results of Model 8.3 in more detail, it is possible to arrive at some interesting conclusions concerning the effects of positive and negative externalities and the interaction between them. The coefficients for Parks and Heavy Industry have been standardised in order that they measure the effect of one kilometre squared of the landuse, at one kilometre distance from a property. For instance, the coefficient for Heavy Industry is -24960. This means that a heavy industrial area, one kilometre squared in size, will decrease the price of a property located one kilometre away by £24,960. A similar explanation applies to the Parks coefficient. However, it would be unwise to extrapolate too much beyond the range of values that exist for the areas of each landuse in Cardiff. For heavy industrial sites, this is within the range of 0.165 - 0.358 kilometres squared, whilst for parks, the range falls within 0.0062 - 0.130 kilometres squared. For a property located one kilometre away from these landuses, the estimated effects on price for each of these ranges would be £1150 - £4260 and £6 - £2691 respectively. Imposing such parameters will prevent excessive predictions being made using the model. Figures 8.1 & 8.2 illustrate the estimated distance decay curves for heavy industrial sites and parks and open space. It can be seen how the externality effects decrease with increasing distance from the source, and how they tend towards convergence. Both externalities have the greatest effects up to 0.25 kilometres from the landuse, suggesting that visible presence may be important. Figure 8.2 implies that the majority of Parks only have a slight influence upon property prices in the immediate proximity, although this is a reflection of their size. The coefficients for the remaining externalities (Bute Park, Primary Schools and Sports centres) represent the effect on price of properties located one kilometre away. The effect of Bute Park is summarised in Figure 8.3 and has a gentle estimated distance decay curve, as the β-value suggests. Conversely, the effect of proximity to Primary Schools and Sports Centres are much more localised, and have a negligible effect a short distance away.

The geographies of the three main externalities (Bute Park, Heavy Industry and Parks) are illustrated in price surfaces in Figures 8.4 - 8.6 respectively. These were generated with the *GRID* toolbox in ARC/INFO using estimated parameters in Model 8.3 to calculate the theoretical impact of each externality upon the properties in the Inner Area property

Figure 8.4

Bute Park Price Surface Model 8.3

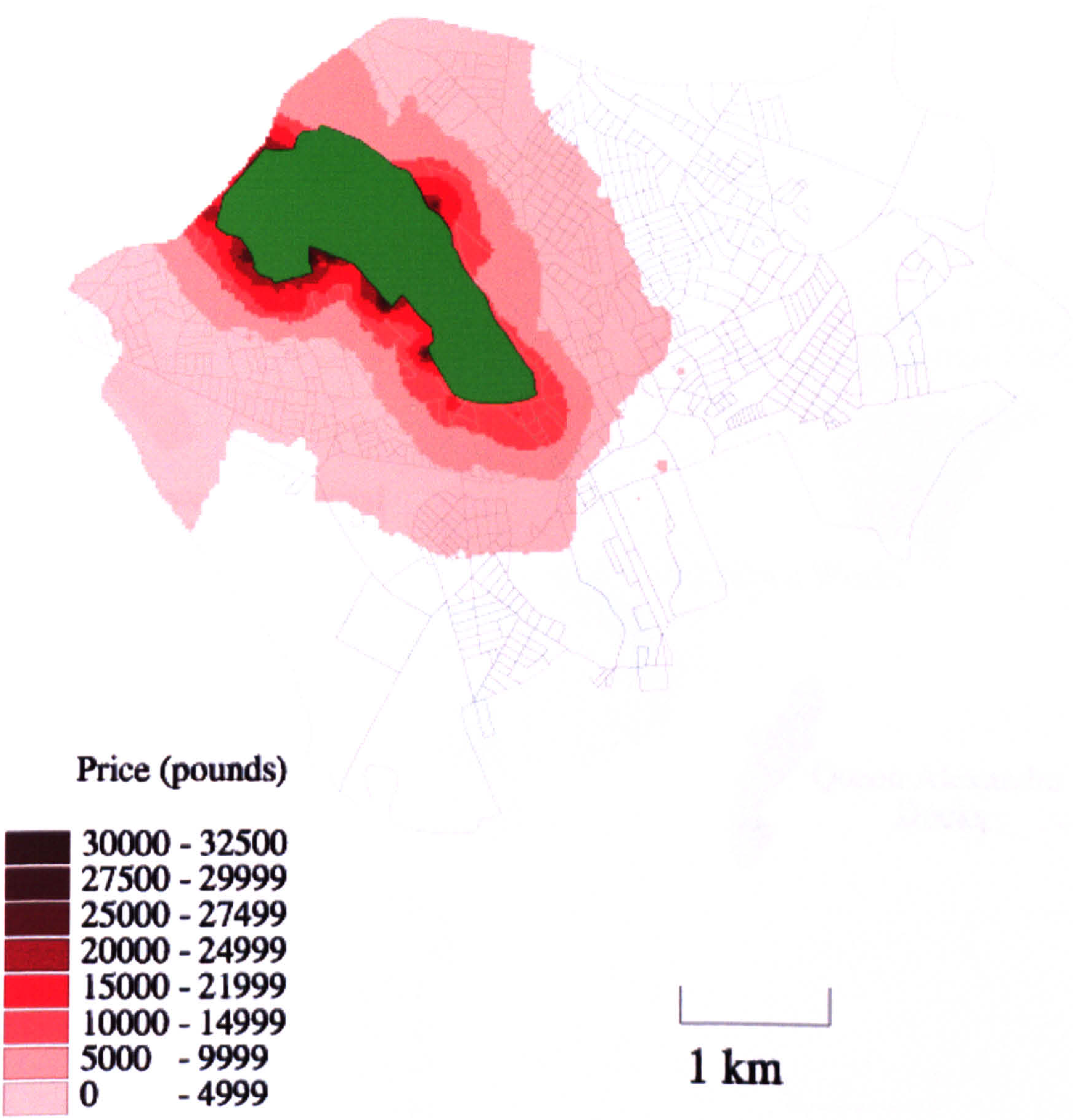


Figure 8.5

Heavy Industry Price Surface Model 8.3

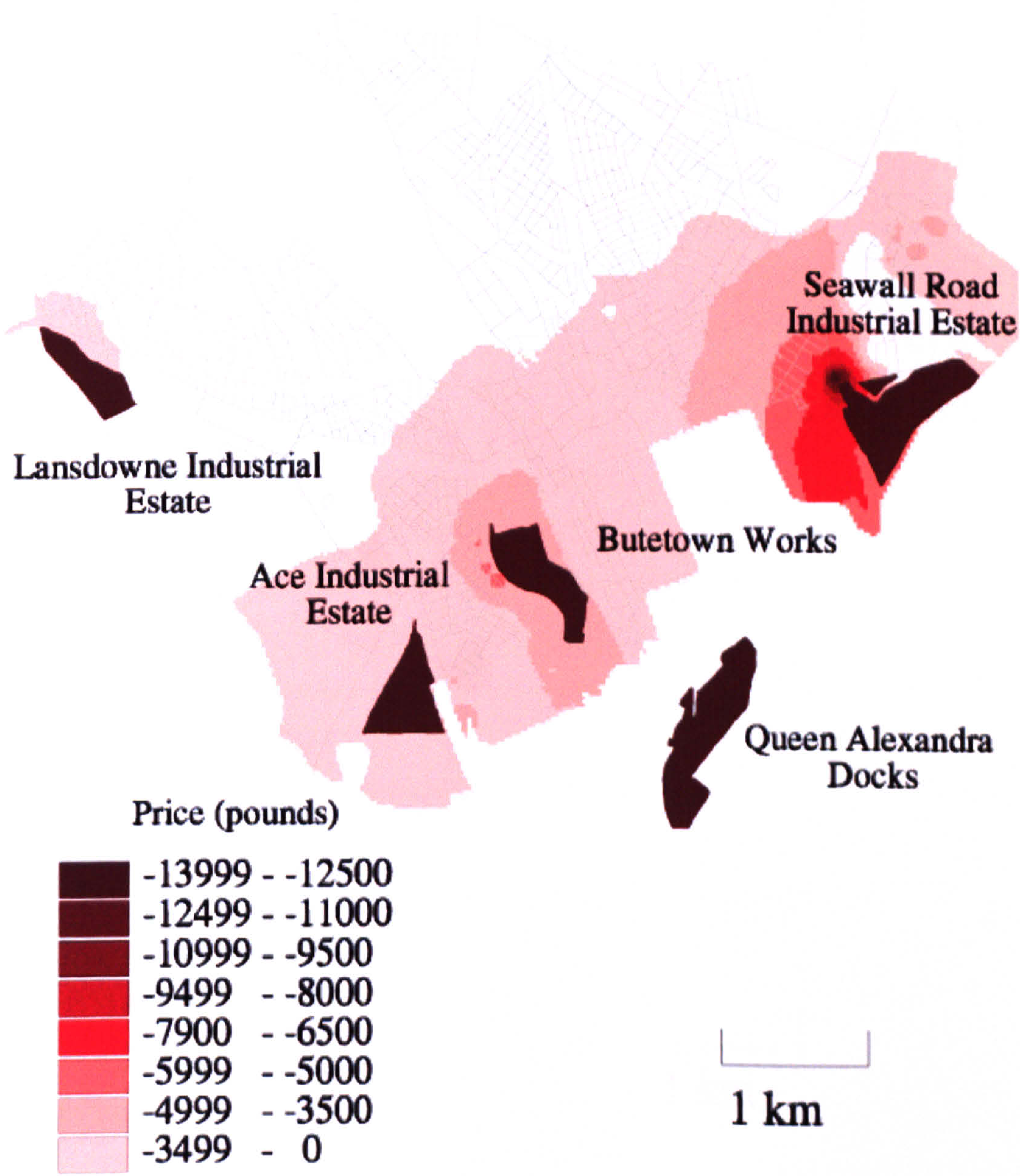
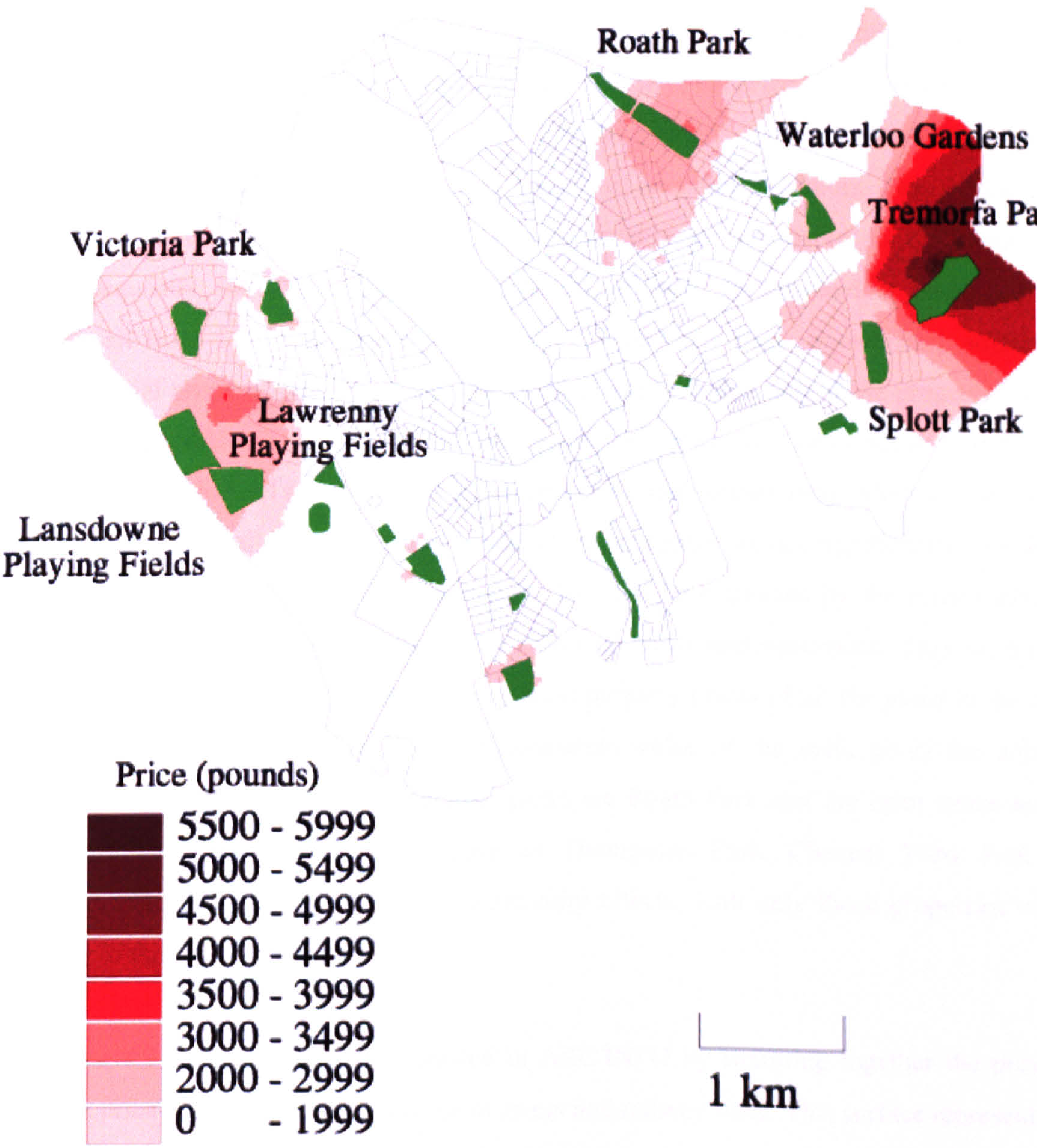


Figure 8.6

Parks and Open Space Surface Model 8.3



coverage. Close inspection of these price surfaces will reveal little pockets of anomalous high and low values. These reflect both the random sample of houses in the Inner Area, and the vagaries of the price surface generating functions in *GRID*. However, since their impact upon the price surfaces are minimal, they can be regarded as white noise, and as such ignored.

Figure 8.4 reveals the importance of Bute Park on property prices in the immediate vicinity, with those next to the Park costing an extra £30 000. This additional premium decays rapidly, and has halved to around £15 000 within a few streets distance. The price surface shows the extent of Bute Parks influence, and how this is affected by the road network and community boundaries. With respect to Heavy Industry, the price surface is more restricted (Figure 8.5). Interestingly, the site which has the greatest estimated effect in Figure 8.1, Queen Alexander Docks, has the least impact in geographical terms. This is due to its isolated nature, away from immediate residential areas compared to the other sites. Consequently, the sites which has the greatest influence on surrounding property prices, Seawall Road Industrial Estate and Butetown Works, are those immediately adjacent to residential property. Moreover, the price surface suggests that only those properties within visible or audible distance of the sites are significantly affected. This contrasts to Bute Park, which has additional amenity value for properties located further away. There is also an area adjacent to Seawall Road Industrial Estate where properties are not significantly affected by their proximity to the negative externality. This can be explained by the price surface in Figure 8.6, which maps the externality effect of parks and openspace. This shows that Tremorfa Park has the greatest influence upon property prices of all the parks in the Inner Area. This is probably due to the compensatory value of the park, given the adjacent negative externalities. Other influential parks are Roath Park and the open space around Llansdowne. The smaller parks, such as Thompsons Park, Channel View Park and Sevenoaks Park have much smaller externality effects, with only those properties with a view of the parks benefiting.

Figure 8.7 is a price surface generated in ARC/INFO by summing together the previous three price surfaces and the influence of rivers and railway lines. This surface represents the cumulative effect of the interaction between the positive externalities in blue and the negative externalities in red. This clearly shows a north-west / south-east split, with positive externalities dominating properties in the former and negative externalities dominating property prices in the latter. Grey areas show where these major externality effects have

Figure 8.7

Spatially Uniform Externality Price Surface

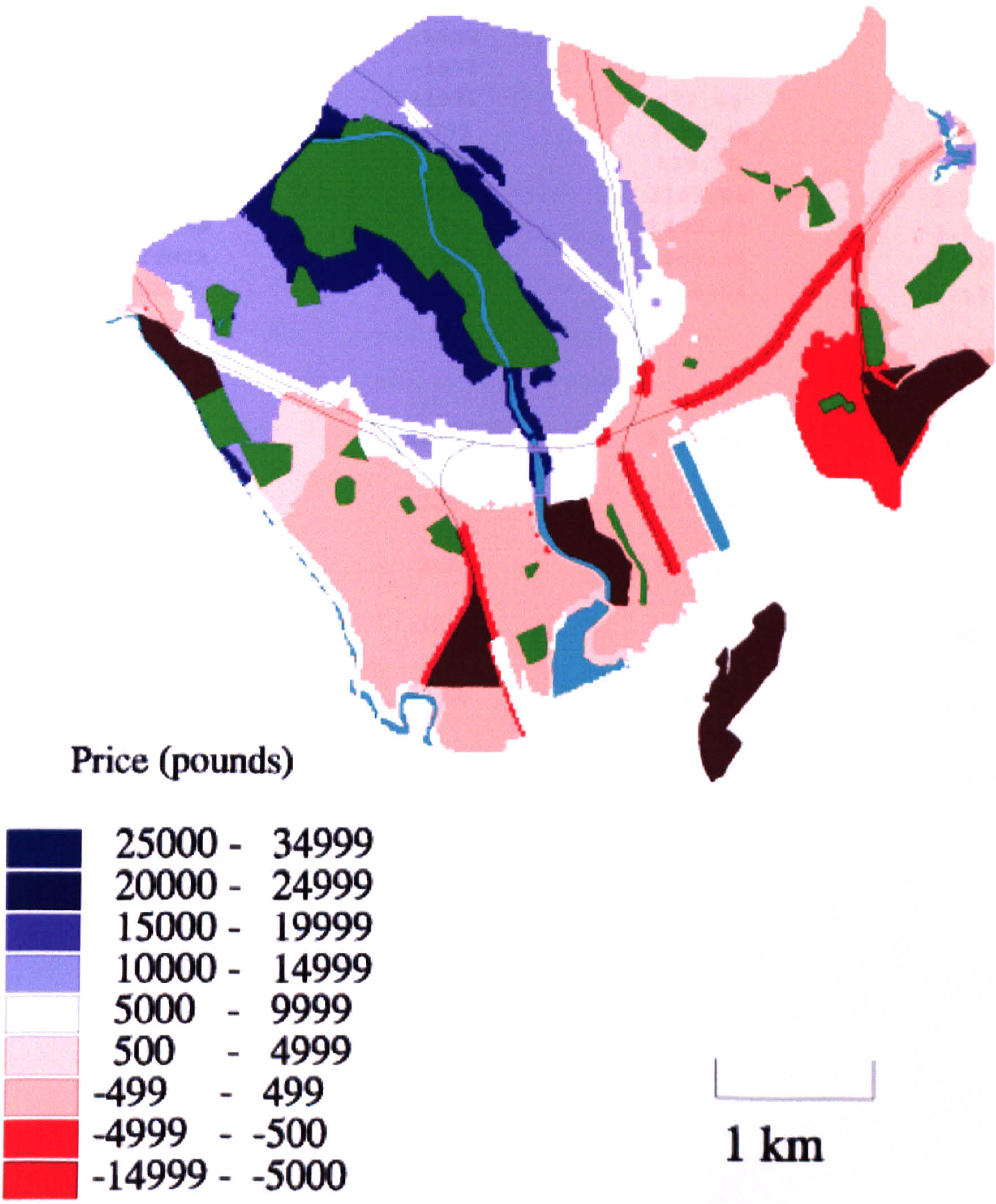


Table 8.8

Model 8.4 - Street Level Locational Attributes

FIXED

PARAMETER	Coeffiecent	S.Error	T-stat
PRIMARY ROAD	-2018	1382.27	-1.46
SECONDARY ROAD	-1463	1153.63	-1.27
RESIDENTIAL ROAD	-504	1073.62	-0.47
POOR 0-50M	-5256	1761.38	-2.98
BELOW AVE 0-50M	-5591	1441.17	-3.88
ABOVE AVE 0-50M	-3227	1357.62	-2.38
POOR 50-100M	-3724	1382.84	-2.69
BELOW AVE 50-100M	-2041	653.30	-3.12
ABOVE AVE 50-100M	-2582	1911.85	-1.35
POOR 100-200M	-1553	1063.70	-1.46
BELOW AVE 100-200M	-1271	672.49	-1.89
ABOVE AVE 100-200M	-793	1086.30	-0.73
NON-RES BUILDINGS	-1521	538.56	-2.82
SCH: WILLOWS	-6689	3169.08	-2.11
SCH: FITZALAN	-7064	1800.68	-3.92
SCH: CANTONIA	-4332	2563.56	-1.69
SCH: CATHAYS	-3481	3155.04	-1.10

RANDOM

PARAMTER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	724	354.86	2.04
HCS Area Level			
CONSTANT	262	105.99	2.47
Sub-Street Level			
CONSTANT	389	156.90	2.48
Property Level			
CONSTANT	1503	150.90	9.96

-2*(log-likelihood) = -530.331

very little influence upon property prices, and these tend to be located on the edge of the Inner Area. The impact of the railway lines and the river can be clearly seen, especially around the docks. The price surface map gives a good impression of the positive impact of Bute Park on property prices in the Inner Area. It also demonstrates the complexity of externality effects on a local scale, with positive and negative areas juxtaposed. A graphic example of this occurs in Splott, with the distinct split between the properties adjacent to Tremorfa Park, and those next to the industrial estate.

8.2.4 Model 8.4 The Street Level Locational Attributes Model

The effects of sub-street level locational attributes are estimated in Model 8.4 (Table 8.8). These can be summarised as the effects of street quality and the secondary school catchment areas. It can be seen that the road type variables are all insignificant, implying that the effects of traffic are uninfluential, although this may be due to the small number of properties sampled on primary and secondary roads, and the fact that the street quality variables also capture the influence of traffic. The effects of street quality are interesting. The omitted dummy variables represents 'good' street quality at various distances from the property. Hence, the immediate street quality (up to fifty metres either side of the property - typically an entire street) has a significant impact upon property price and is illustrated in Figure 8.8. This demonstrates that 'poor' and 'below average' street quality has the affect of reducing the property price by around £5500, compared to 'good' street quality. Interestingly, 'below average' street quality has a more detrimental effect than 'poor' street quality, although this difference is only very marginal. Street quality beyond the immediate property (50-100m and 100-200m) represents the externality effects of the surrounding streets. Model 8.4 indicates that 'below average' and 'poor' street quality within 50-100m of a property are a significant influence upon price, but this is not the case for 'above average' street quality. Street quality beyond 100 metres of the property is generally uninfluential. These results taken together suggest that only when the street quality is below average does it affect price beyond the immediate vicinity of the property (beyond 50m), whilst the distance decay of this effect is steep, and becomes negligible typically within one and half street lengths away from the property. The effect of non-residential buildings in the street also has a significant negative effect upon property, reducing the price by an average of £1500.

Figure 8.8

Valuing the Relative Costs of Street Quality: Model 8.4

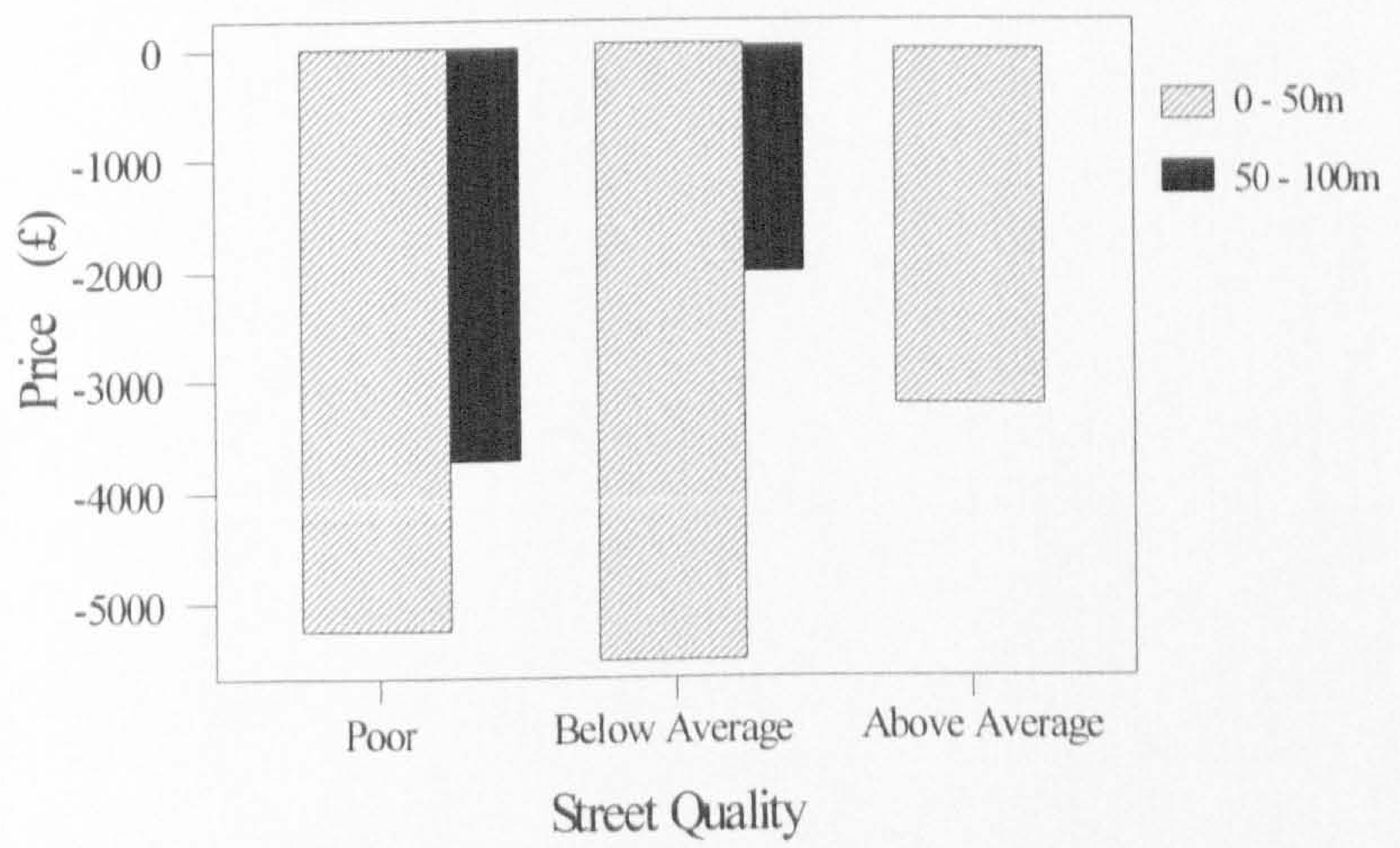


Figure 8.9

HCS Area and Street Level Variances as a Function of Floor Area

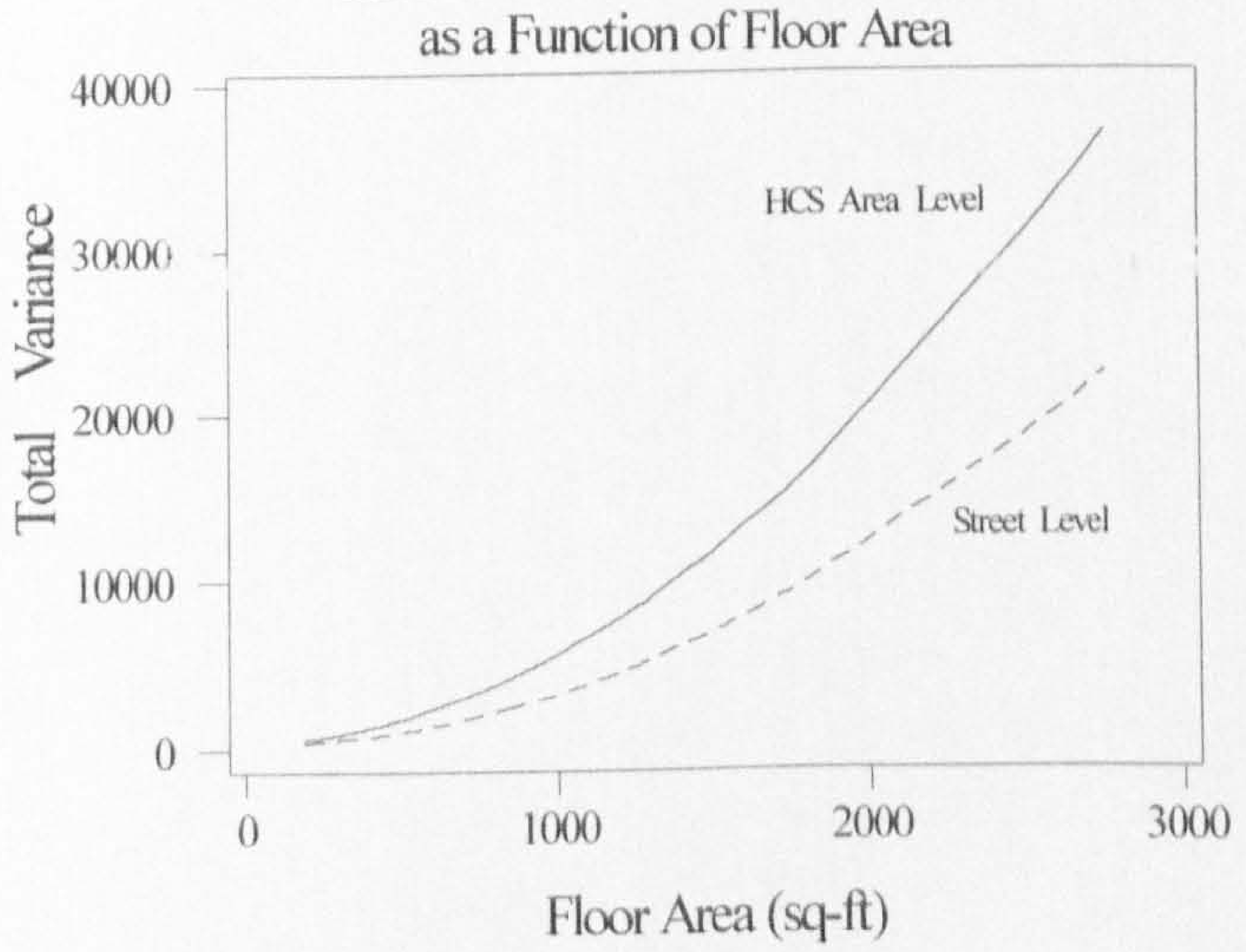


Table 8.9

Model 8.5 - HCS Area Level Locational Attributes

FIXED

PARAMETER	Coefficient	S.Error	T-stat
DENSITY	-0.803	0.31	-2.57
% OPEN SPACE	7073.66	8492.25	0.83
% NON-RESIDENTIAL	-114.32	4948.67	-0.02
Q.SHOP	771	1391.26	0.55
Q.TRANSPORT	2012	2837.76	0.71
Q.SPORT	-87	1014.66	-0.09
Q.PARKS	-1055	1170.97	-0.90
Q.COMMUNITY	713	939.26	0.76
LA > 50%	-2684	1926.37	-1.39

RANDOM

PARAMTER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	631	310.66	2.03
HCS Area Level			
CONSTANT	207	96.072	2.15
Sub-Street Level			
CONSTANT	389	156.90	2.48
Property Level			
CONSTANT	1503	150.90	9.96

-2*(log-likelihood) = -540.648

An important result is the influence that living in a specific secondary school catchment area has on property prices. *Chapter Three* discussed the fact that it is still common for schools to base their intake upon catchment areas, and hence it may be desirable to live within the catchment area of a school with a good reputation. *Chapter Five* described the construction of catchment areas using ARC/INFO, based upon the five schools in and bordering the Inner Area that base their intake upon defined areas. These were then used as dummy variables in the model. Model 8.4 indicates that, relative to the omitted St Teilo High School, only two secondary schools were significantly different: Willows High School and Fitzalan High School. These both had a detrimental affect upon price of around £7000. Both these schools had poor GCSE results and attendance records compared to the omitted school (see Table 5.6). However, the size of these estimates casts doubt upon the accuracy of the coefficients, and as such, is discussed in a later section. Nevertheless, the model does suggest that school catchment areas may be important in some cases.

The random terms indicate that sub-street level locational attributes reduce sub-street level and HCS area level variances by over a half. The property level fixed and random terms remain unchanged since the sub-street level locational attributes do not vary at this level. The difference in the likelihood statistics between this model and Model 8.3 (124) is significant at the 99% level for 10 degrees of freedom, indicating that the Street Level locational attributes significantly explain the variation in house price.

8.2.5 Model 8.5 The HCS Area Level Locational Attributes Model

Table 8.9 summarises the results of adding HCS Area Level locational attributes to Model 8.4. The majority of these variables were constructed using data from the CHCS, and represent access to amenities. The remainder were constructed using ARC/INFO. However, the t-statistics indicate that the only variable that significantly explains the variation in house price is housing density. The remainder are insignificant at the 95% level. This can be explained in part by the lack of detail in the CHCS response data. The variables represent coded responses to questions concerning the quality of local amenities (poor, average, good), aggregated up to HCS Area Level (see *Chapter Five*). Moreover, for the majority of the HCS areas, the types of amenities in question are generally well provided and subsequently their variation will not be of particular importance to the householder. The lack of significance of HCS areas in which the majority of tenure is Local Authority owned can be attributed to the fact that these represent only 5 of 81 HCS areas, and thus are of very

Table 8.10

Model 8.6 - Community Level Locational Attributes

FIXED

PARAMETER	Coeffient	S.Error	T-stat
SOCIAL	1415	599.75	2.36

RANDOM

PARAMETER	Coeffient	S.Error	T-stat
Community Level			
CONSTANT	241	110.83	2.18
HCS Area Level			
CONSTANT	207	96.072	2.15
Sub-Street Level			
CONSTANT	389	156.90	2.48
Property Level			
CONSTANT	1503	150.90	9.96

-2*(log-likelihood) = -554.892

little significance. Therefore, at the HCS Area Level, the most significant factor influencing property prices are the number of houses per square kilometre. High density areas have a negative effect upon property prices compared to lower density areas on the edge of the Inner Area.

The random term suggests that this has a marginal effect upon reducing HCS area level and community level variance. However, the difference in the likelihood statistics (10.32) by the addition of this variable is significant at the 99% level with 1 degree of freedom indicating that housing density is a significant factor in explaining property price variation, however small.

8.2.6 Model 8.6: The Community Level Locational Attributes Model

Table 8.10 summarises the addition of community level locational attributes. These have been basically reduced to the measure of social class used in *Chapter Seven*. Similar to the results in *Chapter Seven*, social class was significant, albeit to a lesser extent, whilst the magnitude of the coefficient was almost half of that in Model 7.12. This indicates that previously, the social class variable had been acting as a proxy for unaccounted locational attributes. With respect to the random terms, the addition of the social class variable has reduced the variance at the community level by 60%, whilst the difference in the likelihood statistics (14.2) indicate that it significantly explains house price variation at the 99% level with 1 degree of freedom. An examination of the variance at each level suggests that the majority of the unexplained variation now occurs at the level of the individual property and the least at the HCS area.

8.2.7 Conclusion

The above section has explored how locational attributes influence property prices, and it can be seen that their effect depends upon the level of resolution that they operate. Generally, the most influential locational attributes operate at the property level, whilst the least influential operate at the community level. Moreover, the results have demonstrated the distance decay nature of many of the locational attributes, particularly those at the property level. It has been shown that this depends upon both proximity to, and the magnitude of the externality, and that positive externalities, such as parks, have a greater impact than negative externalities of comparable size. The analysis also demonstrated that

Table 8.11
Model 8.7 - HCS Area Level Floor Area Interactions

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	203	77.39	2.62
HCS Area Level			
CONSTANT	240	100.99	2.37
FLOOR AREA	0.0052	0.0016	3.32
FLOOR AREA / CONSTANT	0.82	0.30	2.74
Sub-Street Level			
CONSTANT	408	143.21	2.85
Property Level			
CONSTANT	1312	134.89	9.73

-2*(log-likelihood) = -580.658

Table 8.12

Model 8.8 - Street Level Floor Area Interactions

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	186	73.63	2.52
HCS Area Level			
CONSTANT	245	99.29	2.47
FLOOR AREA	0.0046	0.0016	2.82
FLOOR AREA / CONSTANT	0.96	0.31	3.066
Sub-Street Level			
CONSTANT	343	145.18	2.36
FLOOR AREA	0.0027	0.00138	1.94
FLOOR AREA / CONSTANT	-0.25	0.26	-0.96
Property Level			
CONSTANT	1211	129.88	9.33

-2*(log-likelihood) = -588.011

street quality experiences a similar distance decay, and that neighbouring street quality is important, particularly if it is poor. The next section continues this analysis by exploring the spatial variation of locational externalities. In particular, the interaction between structural attributes and locational attributes shall be examined.

Section 8.3 Housing Submarkets and Spatial Parameter Drift

8.3.1 Introduction

In the previous models, the housing market was conceived as operating as a unified whole and so the attribute prices remained constant across the Inner Area. *Chapter Seven* demonstrated that this was not the case for the whole of Cardiff, with submarket conditions causing the implicit price of floor area and social class to vary at higher levels. Hence, it can be hypothesized that the Inner Area housing market will operate under similar conditions. The aim of this section is to ascertain the extent to which submarket conditions influence the valuation of locational externalities. In particular, whether the value of particular externalities are more in some areas than others, and how this relates to the housing stock in these areas. Before this can be achieved, however, the spatial variation in the implicit prices of the structural attributes needs to be accounted for.

8.3.2 Spatial Variation in Structural Attribute Values

Chapter Seven concluded that the implicit price of floor area varied with community context. To capture this effect, Model 8.6 was re-estimated allowing floor area to vary at the community level. However, the resulting floor area random terms were insignificant. As a result, the model was re-estimated allowing floor area to vary at the HCS area level, producing significant results - see Table 8.11. The random part of Model 8.7 suggests that the unit price of floor area varies between HCS areas, with the price per square foot being more expensive in HCS areas with higher than average house prices. This departure from *Chapter Seven*, where submarkets were seen to operate at a larger scale, may be explained by the heterogeneous nature of the Inner Area compared to the Cardiff housing market as a whole. Since the housing stock changes across much smaller distances, it may be expected that supply and demand conditions will also vary at this scale.

Table 8.13

Model 8.9 - Re-estimated Street Level Floor Area Interactions

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	191	75.78	2.52
HCS Area Level			
CONSTANT	259	102.064	2.54
FLOOR AREA	0.00449	0.00164	2.73
FLOOR AREA / CONSTANT	0.875	0.308	2.84
Sub-Street Level			
CONSTANT	335	143.66	2.33
FLOOR AREA	0.0029	0.00144	1.99
Property Level			
CONSTANT	1210	129.97	9.31

-2*(log-likelihood) = -587.278

Table 8.15

Model 8.10 - Street Level Park Interactions

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	163	63.44	2.56
HCS Area Level			
CONSTANT	270	95.71	2.82
FLOOR AREA	0.0044	0.0013	3.30
FLOOR AREA / CONSTANT	0.67	0.26	2.55
Sub-Street Level			
CONSTANT	389	132.29	2.94
FLOOR AREA	0.0006	0.000625	0.96
PARKS	32798	131192	0.25
PARKS / CONSTANT	-5.073	0.84	-6.05
PARKS / FLOOR AREA	-0.0093	0.0037	-2.54
Property Level			
CONSTANT	1347	128.45	10.49

-2*(log-likelihood) = -587.631

Despite the fact that submarket conditions probably would not operate at such a small scale, floor area was also allowed to vary at the sub-street level. The random part of Model 8.8 (Table 8.12) showed that the co-variance between floor area and average house price did not significantly vary at the sub-street level, as was expected, but the unit cost of floor area did. This was confirmed when the model was re-estimated without the covariance term (Model 8.9 - Table 8.13), with the likelihood ratio indicating that the floor area variance term was significant at the 99% level. This unexpected result is discussed in the next sub-section. Figure 8.9 shows how both street and HCS area variances vary as a function of floor area. The total sub-street level variance is roughly half that of the HCS area, whilst the relationship between house size and house price variation is much gentler

The variation of floor area at the street and HCS area levels has changed some of the structural attribute estimates. Table 8.14 shows that the implicit price of floor area in detached housing has halved, whilst the variable measuring the effect of mid-terraces with two bathrooms has become insignificant. This implies that both these variables had been capturing the spatial variation of the floor area coefficient, compensating for the underestimation of the price per square foot in larger properties.

Table 8.14

Structural Attributes in Model 8.9

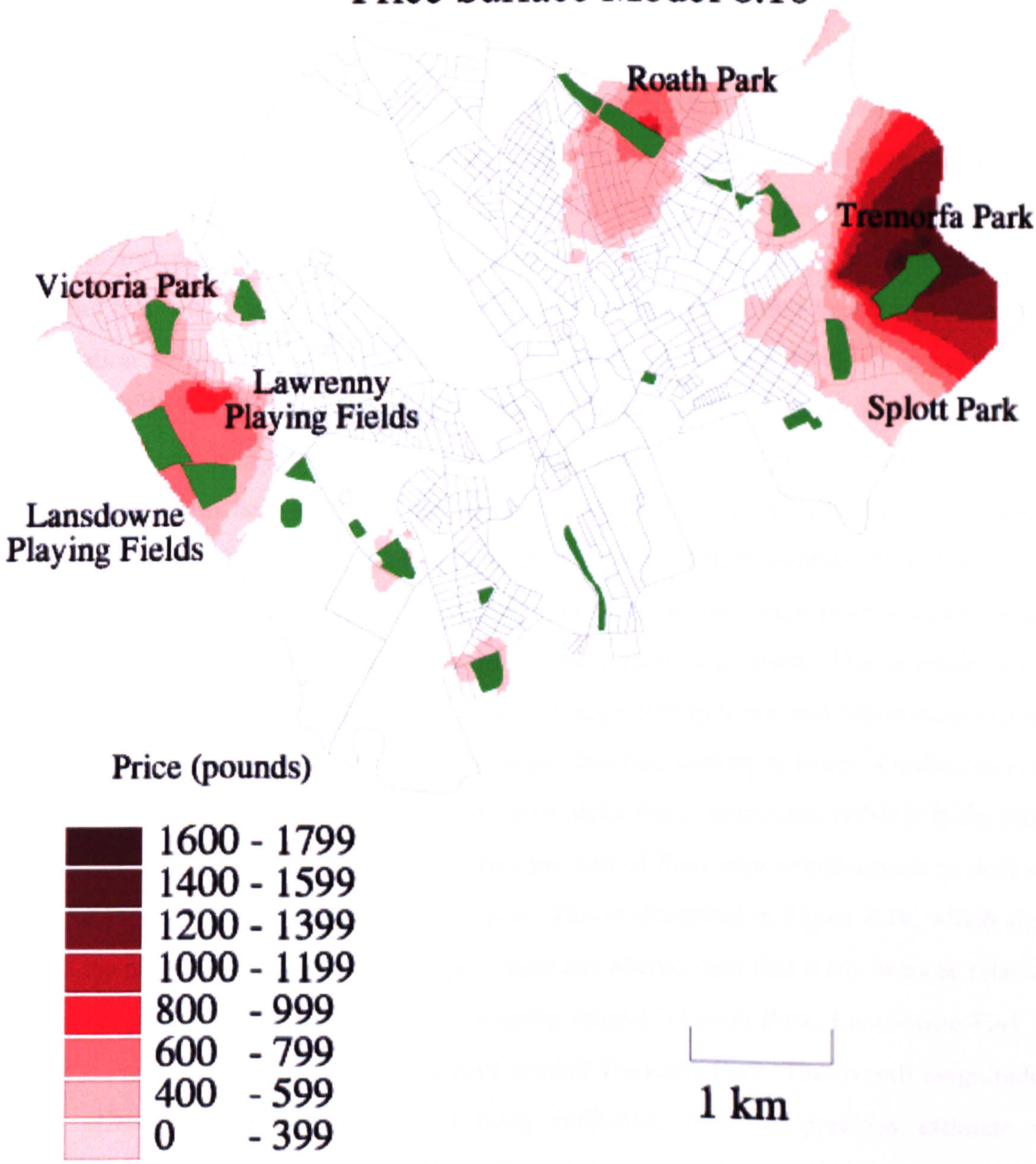
PARAMETER	Coefficient	S.Error	T-stat
CONSTANT	44672	240.26	185.93
FLOOR AREA	30.96	1.95	15.86
FLOOR D	4.58	2.22	2.06
MT BATH 2	3987	2901.78	1.37
FULL CH	3346	724.98	4.62
GARAGE	3938	790.56	4.98
ORP	2677	877.63	3.05
GDN:NONE	-5131	1013.94	-5.06
GDN:5-50M	1980	785.26	2.52
NEEDS MODS	-4761	961.11	-4.95

8.3.3 Spatial Variation in Locational Externalities

The above model has captured the spatial variation of the implicit prices of structural attributes. This sub-section will explore how property level locational externalities vary across the Inner Area. These have been shown to have the greatest impact upon property

Figure 8.10

Parks and Open Space
Price Surface Model 8.10



prices of all the locational attributes, and hence are more likely to have a differential effect across the housing market. Since the externalities are more likely to vary on a street by street basis, rather than on a HCS area or community basis, the variance and co-variance terms of the different externalities were added to level two of Model 8.9. The differential impacts of each externality can then be assessed and new prices surfaces generated that take into account both the sub-street level variation of the externality and the differential effect of floor area.

8.3.3.1 Parks and Open Space

Table 8.15 summarises the results of the random part of Model 8.10, which models park and open space externalities. The first thing to note is the insignificance of both the floor area variance term and the Parks variance term. Their insignificance implies that their implicit prices are constant between streets within a given HCS area. Instead, it is the co-variance terms that are of interest. The likelihood ratio statistic states that the additional two co-variation terms at the sub-street level are significant at the 95% level, when compared to Model 8.7. The negative co-variance between parks and the constant term suggests that marginal impact of parks decreases as average street property price increases, whilst the covariance term between parks and floor area implies that parks also have a marginally bigger impact upon streets which have smaller housing. Taken together, Model 8.10 states that parks have greater impact on property prices in streets which have smaller, cheaper property, than in streets which have larger, more expensive property. This is intuitive since smaller, cheaper property tends to be located in high density areas, and hence access to open space will be more valued than in areas where housing density is lower. Furthermore, this result implies that it is not the implicit price of parks that is important, rather it is the impact it has on the structural attributes. The price per unit of floor area would appear to drift with respect to proximity to parks and open space. This is illustrated in Figure 8.10, which shows that the price surface for parks and open space has altered such that it has become relatively more expensive in the higher density housing around Victoria Park, Lansdowne Park and Roath Park, and relatively less expensive around Tremorfa Park. The overall magnitude of the effect has also declined by a third, indicating that the previous estimate was compensating for the differential effect of house size.

Table 8.16
Model 8.11 - Street Level Heavy Industry Interactions

RANDOM

PARAMETER	Coefficient	S.Error	T-stat
Community Level			
CONSTANT	191	74.61	2.56
HCS Area Level			
CONSTANT	269	103.94	2.59
FLOOR AREA	0.0044	0.0016	2.71
FLOOR AREA / CONSTANT	0.87	0.307	2.83
Sub-Street Level			
CONSTANT	330	143.12	2.31
FLOOR AREA	0.0028	0.001411	1.97
HEAVY IND	8557	18307.77	1.03
HEAVY IND / CONSTANT	1.67	3.37	0.50
HEAVY IND / FLOOR AREA	0.00640	0.0174	0.37
Property Level			
CONSTANT	1216	130.24	9.34

-2*(log-likelihood) = -588.255

Table 8.17
Model 8.12 - Street Level Bute Park Interactions

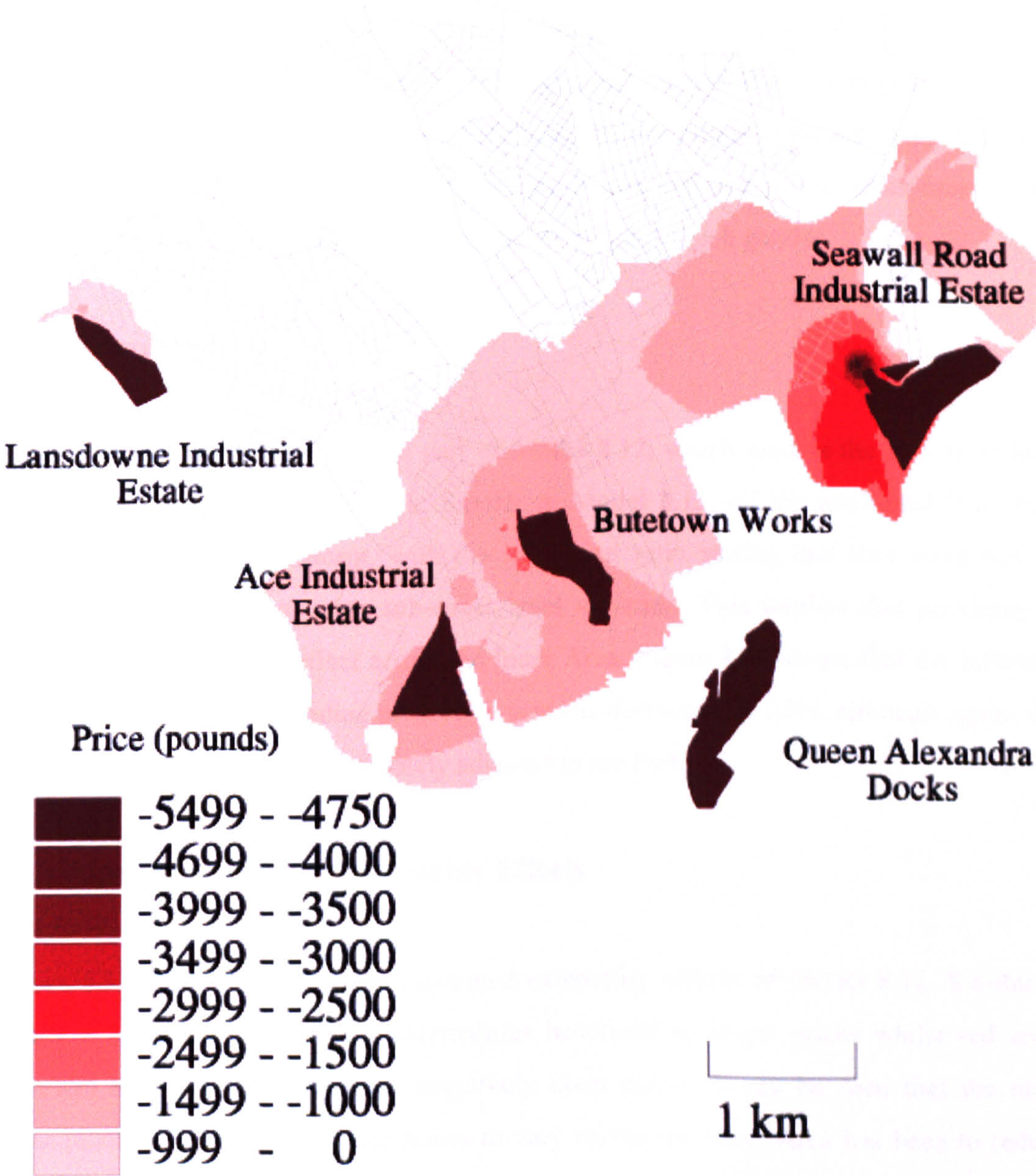
RANDOM

PARAMETER	Coefficients	S.Error	T-stat
Community Level			
CONSTANT	182	72.38	2.52
HCS Area Level			
CONSTANT	221	95.09	2.32
FLOOR AREA	0.0046	0.0017	2.78
FLOOR AREA / CONSTANT	0.911	0.30	3.02
Sub-Street Level			
CONSTANT	348	168.26	2.07
FLOOR AREA	0.0030	0.00147	2.02
BUTE PARK	15613.65	34710.91	0.45
BUTE PARK / CONSTANT	1423.40	1723.35	0.83
BUTE PARK / FLOOR AREA	-1.75	5.14	-0.34
Property Level			
CONSTANT	1191	128.27	9.29

-2*(log-likelihood) = - 590.854

Figure 8.11

Heavy Industry Price Surface Model 8.11



8.3.3.2 Heavy Industrial Sites

Table 8.16 summarises the random part of Model 8.11, which models the sub-street level variation of proximity to heavy industrial sites. In comparison to Model 8.10, all the heavy industry random terms are insignificant, whilst the floor area variance term remains unchanged from Model 8.9. The likelihood ratio states that the addition of the random terms have had no significant effect on explaining sub-street level variation. Hence, it can be concluded that proximity to heavy industrial sites has a constant effect across the Inner Area, regardless of house price and property size. This can be explained by the concentration of such sites in areas of similar housing stock, namely small terraced property. Figure 8.11 is the price surface generated for Model 8.11. This shows that the geography of the externality effect has remained essentially unchanged, which is understandable given the lack of spatial variation. The magnitude of the effect has decreased by around 60%, although most of this decline is concentrated in the areas immediately adjacent to the sites, implying that the externality effect is much gentler.

8.3.3.3 Bute Park

Table 8.17 summarises the random part of Model 8.12, which models the sub-street level variance of proximity to Bute Park. Similar to Model 8.11, all the additional Bute Park random terms are insignificant, with the likelihood ratio stating that they have had no significant effect on explaining sub-street level variation. This implies that proximity to Bute Park has a constant effect across the Inner Area. Figure 8.12 shows that the influence of Bute Park on the surrounding property prices has decreased by 25%, although again, this decline is greatest in areas immediately adjacent to the Park.

8.3.3.4 Combined Spatial Externality Effects

Figure 8.13 is a summary of the combined externality effects of Model 8.12. Similar to Figure 8.7, blue areas represent externalities beneficial to house prices whilst red areas represent externalities that impact negatively upon prices. It can be seen that the main consequence of allowing attribute prices to vary across the Inner Area has been to reduce the impact of the externalities at the extremes and to make their influences more subtle. Hence, the price surface no longer represents a simple mirror image of the externality, with prices simply decreasing with distance. Instead, the spatial variation has created a mosaic of

Figure 8.12

Bute Park Price Surface Model 8.12

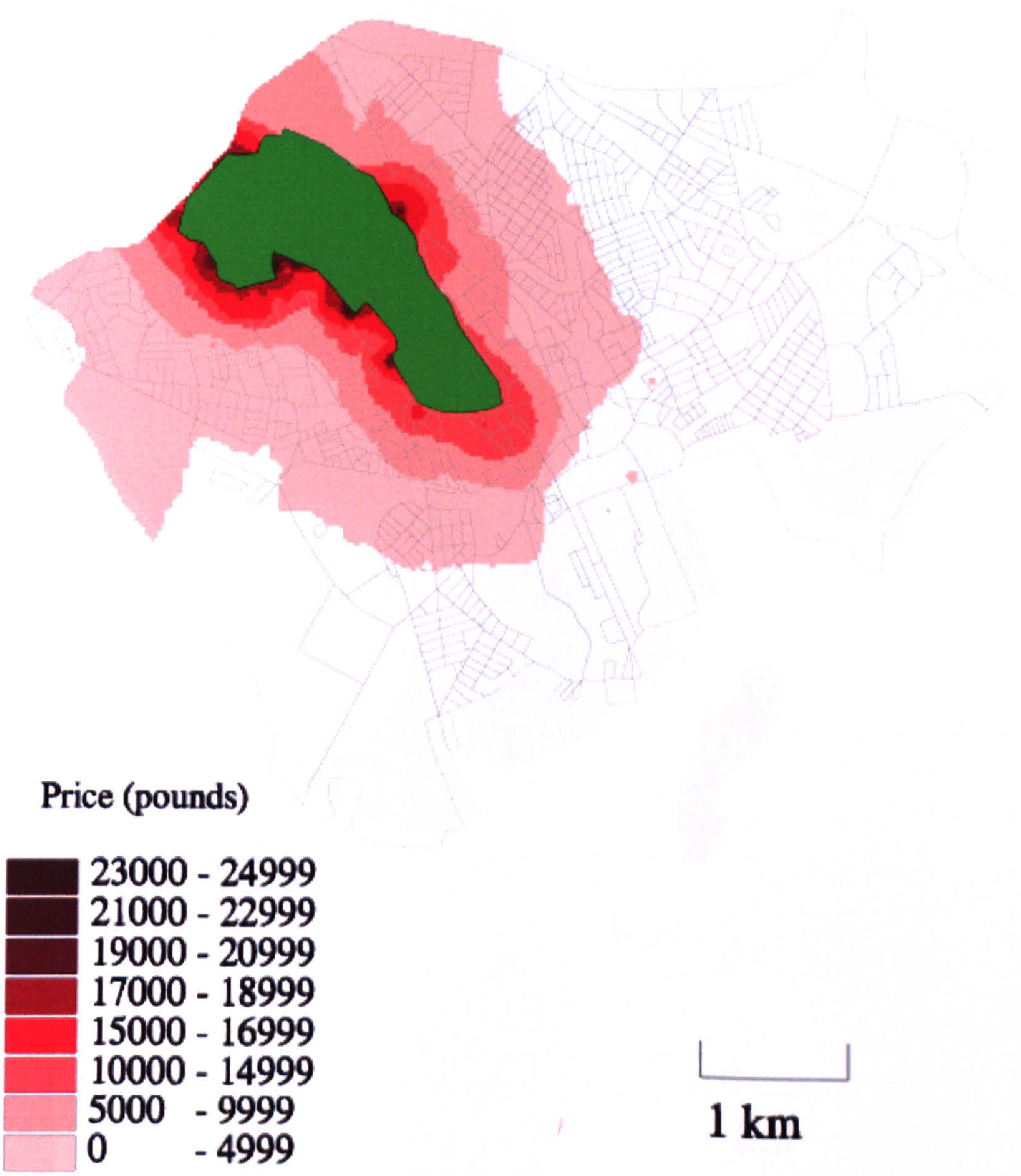
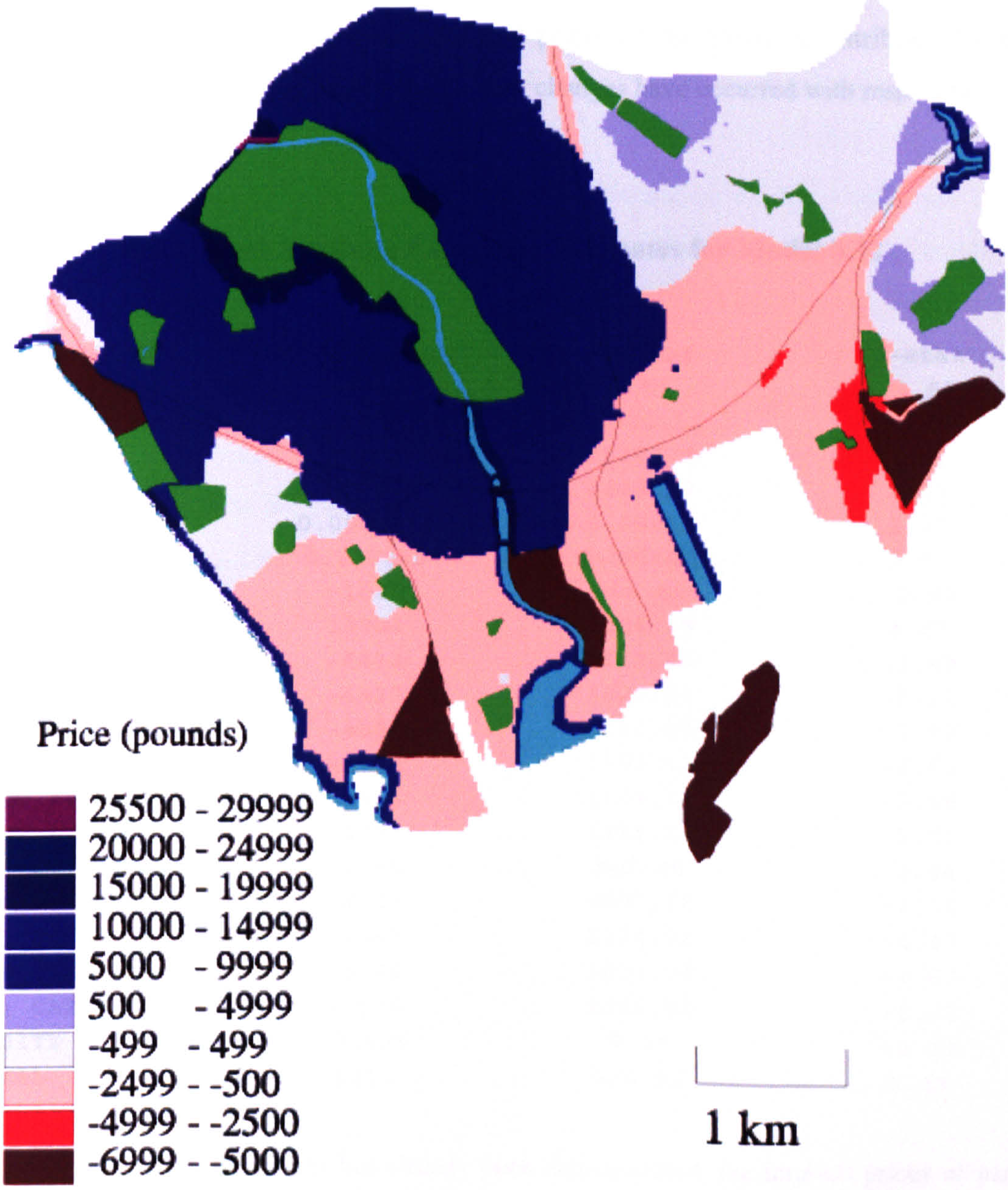


Figure 8.13

Spatially Variant Externality Price Surface



prices, reflecting both strength of the externality and the size of the property. For instance, the influence of Bute Park has declined, such that only the large properties located in Llandaff and Riverside are significantly affected. The areas where overall externality effects are minimal have also increased, and these help highlight localised externality effects such as Roath and Victoria Park. The impact of railway lines have also become negligible in most areas, with the greatest influence upon property prices occurring in Adamsdown and Splott.

8.3.3.5 Implicit Prices of the Locational Attributes

Table 8.18 summarises the estimated implicit prices of the locational attributes for the above price surface. This shows that the greatest changes have occurred with respect to

Table 8.18

Locational Attribute Parameter Estimates for Model 8.12

PARAMETER	Coefficient	S.Error	T-stat
DIST MOTORWAY	-2.23	0.89	-2.51
BUTE PARK($\beta = 0.25$)	8915	14530.11	3.45
PARKS($\beta = 0.25$)	82573	29385.0	2.81
HEAVY IND($\beta = 0.5$)	-9472	4065.23	2.33
SPORTS($\beta = 2.0$)	0.0002573	0.00013	2.02
PRIMARY($\beta = 3$)	0.000045	0.000022	2.01
RAIL 0-50M	-1045	406.60	-2.57
RIVER 0-50M	10704	2504.59	4.27
POOR 0-50M	-6433	1613.51	-3.99
BELOW AVE 0-50M	-6927	1327.89	-5.22
ABOVE AVE 0-50M	-4881	1252.04	-3.90
POOR 50-100M	-5783	1593.11	-3.63
BELOW AVE 50-100M	-5598	1523.15	-3.68
ABOVE AVE 50-100M	-4773	1485.18	-3.21
NON-RES BUILDINGS	-1708	527.06	-3.24
SCH: WILLOWS	-4721	2097.78	-2.25
SCH: FITZALAN	-5968	1334.23	-4.47
SCH: CANTONIA	-3046	1821.00	-1.67
SCH: CATHAYS	-1179	2264.82	-0.52
DENSITY	-0.625	0.29	-2.13
SOCIAL	1317	520.81	2.53

property level externalities. As has already been demonstrated, the implicit prices of parks and open space have almost halved, whilst the implicit price of the heavy industrial area externality has decreased by around 60% to £9400. The influence of Bute Park has also decreased by 25%. Only the externality effects generated by primary schools and sports facilities have remained unaffected, although this is due to the very short distance over which they operate. With respect to sub-street level locational attributes, street quality has become more important, particularly between fifty to one hundred metres from the property, where 'above average' street quality has become significant However, the significance of the school catchment areas have declined, although they still represent an important determinant of property price. At the HCS area level, the importance of housing density has declined now that the differential price of floor area has been accounted for, whilst the effect of social class at the Community level remains substantially unchanged.

Section 8.4 Conclusions

The aim of the chapter was to begin to show how locational attributes influence house prices, and how these vary within the built environment. This was achieved by building a multi-level hedonic model for the Inner Area of Cardiff, and then allowing the locational attributes to vary at different spatial levels. This permitted the locational attributes to enter the house price determination process at the correct level. It also allowed some locational attributes to impact upon property prices to a greater extent than others. This was demonstrated in section two, which showed that locational attributes entering the model at the property level had a greater influence upon property prices than those entering the model at higher levels. Generally, it can be concluded that locational attributes decline in their importance the higher the level at which they operate. Buyers of housing are more concerned with the immediate compositional attributes of the property, as opposed to the more contextual locational attributes of the community.

Section two also examined the way some locational attributes have an externality effect upon property prices, such that their impact depends upon magnitude and proximity. It was shown that different externalities have different impacts on the built environment, and that positive externalities, such as parks, have a wider influence than negative externalities once the size of the effect had been taken into consideration. It was also demonstrated that some externalities, such as primary schools, operate only over very small distance, which implies that proximity in terms of walking distance may be important. Similarly, it would seem that externalities such as rivers and railway lines are only influential if the property is within visual or audible distance respectively. An interesting result from section two was the externality effect attributed to street quality. It was shown that it is not only immediate street quality that impacts upon property price, but also neighbouring street quality appears to be significant. However, the distance decay associated with neighbouring street quality would appear to be greater. The results from section two also inferred that the catchment areas of two secondary schools may be important, although it would appear that these were capturing the unaccounted for differential effects of floor area. Nevertheless, the models suggest that certain school catchment areas do have an impact upon property prices.

An important outcome of this chapter has been the exploration of how the impact of locational externalities varies across the Inner Area, and how they interact with the housing stock. It can be concluded that the implicit price of locational externalities do not vary *per se*, rather the model would appear to indicate that they influence the implicit prices of the structural attributes, and the cost of floor area in particular. As such, they are similar to the effects of social class in the spatial parameter drift specification in *Chapter Seven*, acting as the driving force behind the spatial variation in floor area at the sub-street level. Within the Inner Area, floor area varies between both the HCS area due to supply and demand mechanisms and within the HCS area due to the impact of locational externalities. This results in a complex interaction of structural and locational attributes that was evident in the composite price surface maps, with areas of positive and negative externality effects in juxtaposition. This is especially interesting since it illustrates how externality effects operate over very localised areas. Thus this the chapter has illustrated how locational attributes enter into housing market dynamics, and it can be concluded that, unlike structural attributes, their influence upon property prices are complex and less obvious.

Chapter Nine
Conclusions

Section 9.1 Introduction

The purpose of this final chapter is to draw together the previous eight chapters in a summary of the research. It is divided into four sections. Section two presents a brief review of each chapter, highlighting the main issues raised and the principal conclusions reached. The aim is to present an overview of the entire research, such that the underlying theory, methodology, analysis and results can be clearly seen as a continuous narrative. This overview will feed into section three, which will discuss the implications of the research within a wider context. In particular, the importance of the research with respect to the standard theories of residential location and micro-economics will be evaluated, as well as the role of GIS in spatial analysis and as a means of aiding property valuation procedures. The final sections will discuss how the research might be further developed, and will highlight potential avenues of future research before ultimately concluding the study.

Section 9.2 Synopsis of the Research

Chapter One discussed the underlying themes of the research, and the issues relating to how the built environment is valued. In particular, it put the research into an historical context by discussing how the hedonic house price methodology has grown out of the micro-economic theories of housings market and residential location during the early 1970s. It was apparent that the motivation behind the original hedonic research was to produce empirical verification of the micro-economic theory, principally the estimation of a negative rent gradient from the city centre outwards. As will be explained in section three, the inconsistency of these results was one of the reasons the micro-economic theory fell out of favour.

The main aim of *Chapter Two* was to explore a range of specifications of the hedonic price function. The underlying theme was an investigation of the ways in which space has been incorporated into the specification. It was argued that the traditional hedonic specification was largely aspatial, and ignored the complex spatial structure of the housing market. As such, hedonic modelling using the traditional specification invariably resulted in the estimated parameters suffering from spatial heteroscedasticity and spatial autocorrelation. This implies that the estimates would be biased and inefficient, and therefore would have to be viewed with suspicion. In an attempt to ameliorate this problem, two alternative specifications were developed, that explicitly modelled space; the spatial parameter drift specification and the multi-level specification. The spatial parameter drift specification incorporated space into the fixed part of the model, and can essentially be regarded as the interaction of structural and locational attributes. The multi-level specification incorporated space into the random part of the model, such that housing attributes varied across different spatial levels. The chapter concluded that since most of the previous research had used the traditional hedonic specification, the estimates of how the built environment is valued may have been poorly estimated.

Chapter Three discussed the theory underlying the role of housing attributes, and developed a critique of previous studies. In particular, the role of locational attributes was examined, since historically these had been poorly specified. This included an examination of the concept of locational externalities, particularly with respect to proximity and distance decay functions. Subsequently, the chapter highlighted the mediocre and often quite contradictory results of many previous studies. A specific emphasis was placed upon the measurement of accessibility to the CBD, since the estimation of a negative rent gradient from the city centre outwards was fundamental to both hedonic house price research and the theories of residential location that underpin them. However, it was demonstrated that the empirical evidence for such a rent gradient was a controversial aspect of much of the hedonic house price literature. It was common for many studies to produce either inconsistent or controversial results, such as a positive rent gradient, or find that access to the city centre was insignificant.

The chapter concluded that hedonic house price research could be improved if locational attributes could be more coherently specified. This required not only a means of generating locationally sensitive data, but also a means of handling these data at different levels of resolution. It was therefore argued that a GIS was an ideal medium for hedonic research, and that the spatial analysis capabilities of a GIS might be able to ameliorate the problems of generating locational attribute data that was sensitive enough to capture the heterogeneous nature of the built environment at small scales.

Chapter Four set out the empirical research aims in detail, referring to the discussions and conclusions of the previous chapters. The three main aims were to construct a context-sensitive GIS that could generate and manipulate locationally specific attributes at a high level of disaggregation, to evaluate the three hedonic house price specifications in an attempt to explore the spatial dynamics of the Cardiff housing market, and then to use this information to model locational externalities in an attempt to value the individual elements of the built environment of the Inner Area. The chapter then proceeded to describe the various large and complex socio-economic datasets that were acquired for the research. Three were of particular importance - the house price survey, the Cardiff Housing Condition Survey (CHCS), and Ordnance Survey's ADDRESS-POINT product. The house price survey provided detailed information on the structural attributes and asking price of around 1500 individual properties in the whole of Cardiff, geo-referenced at the level of the unit postcode. The CHCS provided detailed information on the locational attributes of the Inner Area, and in particular, the environmental quality of individual properties and the social characteristics of the resident households. Finally, ADDRESS-POINT allowed both of these datasets to be geo-referenced to resolution of 0.1 metre in the Inner Area. Other datasets, such as the rates register and council tax register, were integral to the matching process and allowed alternative comparisons of property valuations to be made.

Chapter Five described how two GISs were constructed to undertake the research. The Cardiff housing market GIS was formed around three levels of resolution based upon postal and census geographies. Although the level of resolution of the data was not particularly great, it is at a higher level than many of the previous studies and more than sufficient to model the spatial dynamics of the Cardiff housing market. In comparison, a high resolution context-sensitive GIS was constructed for the Inner Area, since a greater level of disaggregation of the data was required. Here, the principal aim was to link the complex socio-economic property datasets to ADDRESS-POINT, such that the housing attributes and housing valuation data were disaggregated to the level of the individual property. The procedures behind this matching highlighted the problems of the integration of address-based socio-economic information, and the inconsistencies of separately compiled datasets. The Inner Area GIS was then enhanced with additional coverages, capturing both the location of non-residential landuses and local amenities in the Inner Area. This GIS was eventually based around five spatial levels: the property, the substreet, the street, the HCS area and the community.

Chapter Six explained how locational attributes were generated for both the Cardiff housing market study and the Inner Area study. Locational attributes for the former were generated from principal components analysis of census data, resulting in the selection of two components. The first component corresponded as a measure of social economic class, whilst the second component corresponded with overall housing quality. In comparison, the Inner Area GIS was used to generate spatially sensitive locational externalities. This involved utilising a whole array of GIS spatial analysis tools to manipulate and analyse the data at various spatial levels. This represented an important part of the research, and as such this study is one of the first hedonic house price studies to have used GIS technology in such a way.

The second part of *Chapter Six* explored the data in an attempt to ascertain the functional forms and relationships between the dependent and independent variables. A range of preliminary hedonic models were subsequently estimated using the traditional hedonic specification, and various diagnostic and statistical tests were applied to evaluate each of the models performance. This exercise in regression model building was undertaken to reduced the risk of specification errors caused by inappropriate functional forms, and the effects of leverage and outlying observations. The chapter concluded by selecting the estimated hedonic model which best described the data, and used this as a basis to subsequent analysis.

Chapter Seven analysed the spatial dynamics of the Cardiff housing market, by exploring how space could be incorporated into the hedonic house price model by using the three specifications discussed in *Chapter Two*. Two methods of analysis were undertaken. Firstly, the ability of each model to capture the spatial structure of the data was measured using diagnostic tests to check for heteroscedasticity and spatial autocorrelation. The second method of analysis was based upon the theoretical consideration that a specification that effectively incorporated space, would produce empirical results that conformed to micro-economic urban theory. This reinforced the arguments in *Chapters Two* and *Three* that previous poor empirical results were due in part to a neglect, or a lack of appreciation of the problems associated with modelling complex spatial data.

More specifically, the analysis attempted to model the form of the hypothesized negative rent gradient from the city centre outwards, and the existence of submarkets caused by disequilibrium in the housing market. The latter is a necessary departure from standard micro-economic theory that emphasizes a perfectly functioning housing market, in which supply and demand mechanisms results in the Pareto Optimum conditions. The identification of submarkets, a basic tenant of hedonic house price theory, contradicts the standard micro-economic theory of housing markets and hence is an important result

With respect to diagnostic tests, the analysis revealed that the hedonic models increasingly explained the spatial structure of the data as the specification changed from the traditional specification, to the spatial parameter drift specification to the multi-level specification. Specifically, the degree of heteroscedasticity and spatial autocorrelation displayed by the model's residuals decreased. It was concluded that the multi-level hedonic specification best modelled the spatial housing market dynamics, although residual heteroscedasticity was still present. This was attributed to both the use of communities in defining submarkets and omitted locational attributes.

In terms of the empirical results, the analysis was successful in modelling the negative rent gradient. It was shown that the rent gradient became progressively more concave as the hedonic models increasingly captured the spatial structure of the data. This is interesting in itself since it suggests that the apparently contradictory results of previous research are probably attributable to a lack of modelling the spatial structures in the data.

The analysis showed that submarkets could be differentiated both by the housing stock, in this case detached housing, and geographical areas (communities). This demonstrates that the housing market is not in equilibrium as the micro-economic theory assumes. Instead, the analysis revealed that the implicit price of the floor area attribute in detached housing was significantly different to the implicit price of floor area in other dwelling types in Cardiff. Moreover, the analysis showed that the implicit price of floor area also differed significantly between communities. Since floor area was shown to be the most important attribute in determining house price, these results are non-trivial. What they suggest is that the housing market does not operate as a unified whole, but rather is segmented into smaller submarkets. The implications that this has for supply and demand mechanisms, and disequilibrium will be discussed in the next section.

Finally, the analysis revealed that the marginal price of structural attributes, and in particular floor area, decreased as social class increased. This indicated that structural attributes become less important in determining house price in areas of higher income households, while locational attributes become marginally more important. This would seem to imply that higher income households place more value on locational attributes, than lower income households. The implications of this are discussed in the next section.

Chapter Eight continued to explore the value of locational attributes, using the multi-level hedonic specification to model Inner Area house prices. Three substantive conclusions were drawn. First, locational attributes influence house prices across different scales of resolution, and the locational attributes that operate at the level of the individual property are more important in determining house price than locational attributes that operate at higher levels. Secondly, the analysis indicated that positive externalities appear to have greater impact than comparable negative externalities, with positive externalities impacting across a wider area. This suggests that buyers of housing may place greater value on proximity to positive externalities, compared to negative externalities within the same vicinity. Finally, the research revealed that locational attributes affect the value of structural attributes, and the marginal price of floor area in particular. This results in a complex interaction between structural and locational attributes, with positive and negative externality effects operating across only very localised areas. This conclusion is of particular interest since it suggests that the value of a house is intimately bound up with its location, and that any changes to the attributes of location may have significant effects upon its value. This is discussed further in the next section.

Section 9.3 Implications of the Research

This section will discuss the implications of the research within a wider context. This can be roughly divided into five identifiable areas; implications for the underlying micro-economic theory, implications for future hedonic house price research, implications for locational externalities and the built environment, implications for GIS and spatial analysis, and the implications for real estate, property valuations and local taxation.

9.3.1 Micro-Economic Theory of Housing Markets

The 1960s and the 1970s saw a proliferation of micro-economic theories and mathematical models of housing markets, residential location and landuse in both the UK and USA (e.g. Muth, 1969, Batty, 1976). These were essentially demand led models that assumed that western capitalist housing markets operated under Pareto Optimum conditions. In such a formulation, the supply and demand of housing were assumed to be in perfect equilibrium. Furthermore, the supply of housing was either ignored or was assumed to effortless adapt to variations in demand. Deductions from these models were used to explain the spatial patterning of residential location, which centred around the concept of a bid-rent function. This function described the process in which households traded-off accessibility to the city centre with housing attributes, and in particular housing size. The result of this trade-off between accessibility and space was a decrease in rents from the city centre outwards, which declined at a decreasing rate.

However, since these models were essentially descriptive, these deductions required verification from empirical evidence. It was explained in *Chapter One* that the motivation behind the early hedonic house price research was to supply this empirical evidence, namely in the estimation of the negative rent gradient, which would strengthen the argument for the concept of the accessibility / housing size bid-rent function. However, as was discussed in length in *Chapter Three*, the estimation of the negative rent gradient has been a cause of controversy in hedonic house price studies, with counter-intuitive or insignificant results being the norm. In addition, subsequent hedonic house price research also questioned the validity of the concept of a market in perfect equilibrium functioning under Pareto Optimum conditions. Instead, hedonic research became increasingly concerned with disequilibrium, with the supply and demand of housing operating under restrictive conditions. Initially, the main emphasis of this research was upon racial segregation in North American cities, and how this affected the housing market. In recent years though, housing market stratification in more general terms has became a predominant concern. However, similar to the negative rent gradient, the evidence for the existence of submarkets was also contradictory. Therefore, both this interest in housing market disequilibrium, and the lack of consistent evidence for a negative rent gradient, contributed to micro-economic housing market theory falling out of favour in the late 1970s.

However, it was explained in *Chapter Two* that the hedonic price function may have been misspecified, whilst *Chapter Three* highlighted the fact that many of the previous studies had used poorly specified data, particularly with respect to locational attributes. Together, these factors may have been responsible for the lack of consistent empirical evidence which is characteristic of much hedonic research . *Chapter Seven* demonstrated that this may have been the case, by modelling the spatial dynamics of the Cardiff housing markets. Using a highly disaggregated database, geo-referenced to a higher resolution than many previous studies, a negative rent gradient was identified. Moreover, this rent gradient became increasingly concave as the hedonic specification became more sensitive to the spatial structures of the data, and hence the spatial dynamics of the housing market. This would seem to verify the concept of a trade-off between access to the city centre and all off the other housing attributes.

Chapter Seven also demonstrated the existence of submarkets, implying housing market disequilibrium. With respect to the separate submarket for detached housing, this can be explained in terms of supply and demand. Detached housing has, on average, the greatest amount of living space of all the dwelling types in Cardiff, and also tend to be restricted to suburban locations, particularly in areas which have good neighbourhood quality. In terms of demand, detaching housing would be desired by larger households, specifically families with children. In addition, high neighbourhood quality would also make detached housing particularly appealing. This interaction between supply and demand would appear to have created specific market conditions, with the more affluent households being able to out bid households on lower incomes.

Submarkets operating within defined geographical areas, in this case communities, can be explained by factors such as imperfect knowledge of the housing market, a desire to live near family and friends, or restrictions placed upon the supply of housing, such that perfect substitution is not possible. In addition, institutions and actors operating within the housing market, such as estate agents, may also reinforce these submarket conditions. It was explained in *Chapter Four* that estate agents use the twenty-six Cardiff communities as a basis for structuring house sales. For instance, houses for sale were grouped into their respective communities, which help to guide potential buyers. More importantly, since estate agents value property using the comparative method (Millington, 1990), communities would become integral to the valuation process. Therefore, it is hardly surprising that property prices reflect community boundaries. This was shown in the spatial autocorrelation maps in *Chapter Seven*, which indicated that house price residuals within a community were more similar than house price residuals between communities.

Therefore, the results of *Chapter Seven* have several implications for the micro-economic theory of housing markets and residential location. The main implication is the estimation of the negative rent gradient which gives support to the concept of the bid-rent function and the trade-off between accessibility to the city centre and house size. However, the implications of housing submarkets would appear to contradict the assumptions of housing market equilibrium that are fundamental to the micro-economic theory. In particular, the research points to the possible importance of estate agents as structuring the valuation process and hence to some extent influencing the spatial dynamics of the housing market.

9.3.2 Implications for Hedonic House Price Research

The hedonic pricing method is very well established, particularly in North America. From its origins as a method of producing empirical evidence to underpin the micro-economic theory of housing markets, it has subsequently been used as a common method of imputing the value of intangible attributes of housing. It has especially been employed to impute environmental benefits, such as the amenity value of forests. Furthermore, the implicit prices estimated by the hedonic price function are regularly used in demand equations to estimate the costs of specific amenities, such as the costs of air and noise pollution. *Chapter Two* explained how substantial intellectual energy had been expended over the past few years on the specification of these demand equations, with comparably very little on the traditional specification of the hedonic price function. It is now clear that this traditional specification may be misspecified with respect to space. The corollary of this is that the estimates of the hedonic price function may be biased and inefficient, and that this inefficiency will follow through to subsequent demand equations. If the results of these demand equations are used subsequently to inform government policy on the costs of cutting urban air pollution, say, then the consequences of this misspecification become non-trivial.

In response to this concern, *Chapter Seven* investigated the ability of three hedonic specifications to model the spatial structures of the housing market. The results of this investigation indicated that the traditional specification produced the most inefficient models with respect to heteroscedasticity and spatial autocorrelation. With respect to this criteria, the best specification was the multi-level specification. Conceptually, housing market dynamics were also best described by the multi-level specification. This specification allowed the spatial structures inherent in the valuation of property, such as community boundaries, to be modelled explicitly, whilst the specification also allowed both the compositional effects of the housing stock and the contextual effects of location to be modelled simultaneously. However, it can be argued that the multi-level specification places a too rigid a criterion upon the delimitation of spatial boundaries within the housing market, and does not allow for possible spill over effects between them. In this respect, the spatial parameter drift specification is conceptually more appealing, although as was shown in *Chapter Seven*, the inability of this specification to model the spatial variation in locational attributes, caused heteroscedasticity in the estimated parameters. Therefore, it can be concluded that the multi-level hedonic specification is the most efficient at modelling housing market, and this should have implications for future research.

9.3.3 Locational Externalities and the Built Environment

The effects of locational externalities have been a common concern in the urban economic literature, and a particularly important concept of how the built environment is valued. It has traditionally been couched in terms of power and conflict, specifically in how negative externalities impact upon peoples' lives and property values, and how people subsequently come together to exclude them from their locality. Although these incidents are perhaps overstated, it is sill the case that in recent years, NIMBY (Not In My Back Yard) -ism has remained an important issue within the UK. This research, and particularly the analysis in *Chapter Eight*, has demonstrated that locational attributes are a very important part of the house price valuation process. Moreover, *Chapter Seven* concluded that higher income households may value locational attributes marginally more than structural attributes, and hence that these may represent a larger proportion of the value of higher priced properties. This was supported by the findings of *Chapter Eight*, that concluded that the value of structural attributes are intimately bound up with locational attributes. If this is the case, then it gives credence to the argument that higher income households are more likely to come together in action to protect their property prices, than lower income households, since locational externalities represent a larger investment in the price of their property.

Chapter Eight also demonstrated the complex geography of locational externality effects at the local level. Positive and negative externalities were shown to be juxtaposed across very small scales, and their effects upon property prices could be measured on a street-by-street basis. This has important implications for hedonic house price research, since much previous work has been undertaken at much lower resolutions. More specifically, any estimation of the value of amenities will have to take these small scale variations into account. This has joint implications with the previous discussions concerning the specification of the hedonic house price function. Both the ability to model spatial effects and the resolution of the data are important if efficient estimates of housing attributes are to be made.

9.3.4 Implications for GIS and Spatial Analysis

Questions concerning the resolution of the data, and the measurement of locational attributes, has connections with the role of the GIS in this research, and within spatial analysis in general. In the past, the role of GIS in spatial analysis has been questionable. In *Chapter Three*, it was discussed that commentators such as Openshaw (1995), have argued that GIS should move away from Geographic Information Handling to Geographic Information Using, and that GIS has been under utilised as a powerful tool in spatial analysis, especially in the social sciences. One of the problems that has contributed to this under utilisation has been a lack in the availability of spatially disaggregated socio-economic data, geo-referenced to a high resolution. *Chapters Four* and *Five* of this research has shown that, with the advent of digital products such as ADDRESS-POINT, and appropriate matching techniques, the problems of geo-referencing socio-economic data to a high resolution are being slowly addressed. Large and complex addressed based datasets, such as the CHCS, rates register and council tax register, are now capable of being linked and manipulated at a level of disaggregation not possible before. Indeed, this research has constructed and utilised a GIS at a level of disaggregation and complexity not used before in any hedonic house price study, with the housing data having been modelled at the appropriate level of resolution for the first time.

Chapter Six described in detail the spatial analysis tools available in ARC / INFO which were used to manipulate this disaggregated data to derive new data at various levels of resolution. These include the traditional spatial analysis tools, such as POINT-IN-POLYGON, as well as more sophisticated tools, such as NETWORK and GRID. These tools were able to generate locationally sensitive data, that took into account the vagaries of urban form, such as the topology of the street network. The research also demonstrated the effectiveness of GIS in the visualisation of data, in particular the model's residuals which demonstrated the importance of communities in structuring the data, and the externality value surfaces which were an important part of the analysis in *Chapter Eight*. It can therefore be concluded that it is the role of the GIS in the research, and the high resolution of the data, which are in no doubt responsible for the quality of the results.

The research also demonstrated the continued importance of postcodes as the basis of matching address-based datasets. Such datasets are becoming increasingly important with the advent of geo-marketing, and private companies such as supermarkets collecting information upon individual people. By highlighting the procedural problems with address-based matching, *Chapter Five* also concluded that the standardisation of addresses should be paramount, and that this is now increasingly possible with advent of the British Standard BS7666.

9.3.5 Implications for Real Estate and Property Valuations

It has already been seen that a GIS is a perfect medium for handling housing attribute and valuation data. Increasingly, it has been used by real estate agents to aid their business, especially in North America (Dixon, 1992). There is therefore a good argument for its introduction into the UK. Property valuation is at best a very inexact science (Millington, 1990), due in part to a lack of available, comparable comprehensive databases. This is quite surprising since the comparative method is the most common technique of valuing domestic property, and this relies in part upon archival data records. The introduction of information technology, and GIS in particular, may go some way to alleviate the present uncertainties in the valuation procedure. This has added significance since local government taxation is now based upon capital valuations of property.

Section 9.4 Future Research and Overall Conclusions

The final part of this chapter will conclude by discussing future possible research. An obvious area of potential research lies in expanding the detailed Inner Area study to cover the entire Cardiff housing market at the same resolution. The implications behind this is that effects of locational externalities may differ in suburban locations, where the housing stock displays greater homogeneity and housing density is generally lower. To be comparable with the Inner Area study, a similar context-sensitive GIS would need to be constructed. This would require ADDRESS-POINT and the digitised street network for the entire Cardiff housing market. Moreover, the fact that the CHCS does not extend beyond the Inner Area implies that extensive field work measuring street and neighbourhood quality would also be required. However, the resulting research would be the most comprehensive study of the effects of locational externalities ever undertaken, and shed light upon the qualitative and quantitative differences between the inner city and suburban valuation of location.

Another potential area of research would be an investigation of the effectiveness of Geographically Weighted Regression upon modelling the spatial dynamics of the housing market. Geographically Weighted Regression is a recent technique, developed over the past couple of years (Brunsdon et al, 1996), and was briefly outlined in *Chapter Two*. Due to its recent innovation, very little research of any description has been undertaken using this technique, and hedonic house price research is of no exception. Therefore, a comparison of an hedonic specification based upon Geographically Weighted Regression with the three specifications investigated in this research will be of potential interest, particularly the ability of Geography Weighted Regression to model space.

A final, future area of research relates to the continued investigation into the geography of revenue raising at the local level. Previous research has investigated the impact of the council tax at the street and HCS area levels. It would be of values to continue this research at the level of the individual property, especially with respect to exploring the relationship between council tax and locational externalities. Previous research has indicated the importance of location, but the impact of specific externalities has yet to be investigated. A knowledge of how locational externalities affect council tax banding will have important implications upon future investments in the built environment.

Therefore, as a final conclusion, it is apparent that this research has achieved the aims that were set out in *Chapter One*. It has investigated how the built environment is valued, particularly at the local level. By doing so, the research has highlighted the problems that can be expected to be faced when modelling large and complex spatial datasets. Furthermore, the research has demonstrated the importance of GIS in structuring and manipulating the data. Indeed, it is the ability of the GIS to handle large and complex socio-ecomonic datasets at a high level of disaggregation, and to generate locationally sensitive externality data using the spatial analysis tools that allowed this research to produce such detailed results at such a small scale. Together with the availability of increasingly sophisticated datasets, this technology has the potential of opening up new avenues of research into the built environment that has just not been possible before. This research is an illustration of what can be achieved. Future research may be able to build on this case study, improving the GIS and spatial analysis techniques to unravel the complexities of the built environment at even finer levels of resolution.

Appendix

Glossary of Abbreviations

Abbreviations Used in the Models

Structural Variables

Floor Area	Total Floor Area (sq-ft)
Ave Bed	Average Bedroom Floor Area (sq-ft)
Ave Rec	Average Recreation Room Floor Area (sq-ft)
Ave Kit	Average Kitchen Floor Area (sq-ft)
ET	End-Terraced Dwelling
MT	Mid-Terraced Dwelling
SD	Semi-Detached Dwelling
D	Detached Dwelling
FCB	Flats in Converted Building
FPB	Purpose Built Flats
M	Maisonette
B	Bungalow
EL	End-Link Dwelling
ML	Mid-Link Dwelling
Beds	Number of Bedrooms
Recs	Number of Recreation rooms
Baths	Number of Bathrooms
Showers	Number of Shower rooms
Full CH	Full Central Heating
Part CH	Partial Central Heating
Gas	Gas Central Heating
Garages	Number of Garages
ORP	Off-Road Parking
New	Age: New
Post 1964	Age: Post 1964
1918-64	Age: 1918 - 1964
Pre-1918	Age: Pre-1918
Gdn: None	Garden: None
Gdn: < 5m	Garden: Less than 5 metres
Gdn: 5-50m	Garden: 5 - 50 metres
Gdn: > 50m	Garden: More than 50 metres
Needs Mods	In need of modernisation
Swm Pool	Swimming Pool
Con	Conservatory

Locational Variables

Dist CBD	Accessibility to CBD
Dist Mway	Accessibility to M4 motorway
Dist Station	Accessibility to railway stations

Hospital	Proximity to hospitals
Sports	Proximity to sports centres
Community	Proximity to community centres
Institutional	Proximity to institutional centres
Shops	Proximity to local shops
Primary	Proximity to primary schools
Secondary	Proximity to secondary schools
Bute Park	Proximity to Bute Park
Parks	Proximity to parks / open space
Light Ind	Proximity to light industrial land-use
Heavy Ind	Proximity to heavy industrial land-use
Rail 0-50m	Rail 0 -50m
Rail 50-100m	Rail 50 - 100m
Rail 100-150m	Rail 100 - 150m
Rail 150-200m	Rail 150 - 200m
River 0-50m	River 0 - 50m
River 50-100m	River 50 - 100m
River 100-150m	River 100 - 150m
River 150-200m	River 150 - 200m
Primary Road	Road Type: Primary
Secondary Road	Road Type: Secondary
Residential Road	Road Type: Residential
Close	Road Type: Cul-de-sac / Close
Poor 0-50m	Street quality 0-50m: Poor
Below Ave 0-50m	Street quality 0-50m: Below Average
Above Ave 0-50m	Street quality 0-50m: Above Average
Good 0-50m	Street quality 0-50m: Good
Poor 50-100m	Street quality 50-100m: Poor
Below Ave 50-100m	Street quality 50-100m: Below Average
Above Ave 50-100m	Street quality 50-100m: Above Average
Good 50-100m	Street quality 50-100m: Good
Poor 100-200m	Street quality 100-200m: Poor
Below Ave 100-200m	Street quality 100-200m: Below Average
Above Ave 100-200m	Street quality 100-200m: Above Average
Good 100-200m	Street quality 100-200m: Good
Non-res Buildings	Street non-residential buildings.
Sch: Willows	Sch Catchment: Willows High School
Sch: Fitzalan	Sch Catchment: Fitzalan High School
Sch: Cantonia	Sch Catchment: Cantonia High School
Sch: Cathays	Sch Catchment: Cathays High School
Sch: St Teilo's	Sch Catchment: St Teilo's High School
La > 50%	Percentage Local Authority tenure
% Open Space	Percentage of open space
% Non-Residential	Percentage of non-residential land-use
Density	Housing density
Q.Shop	Quality of local shops
Q.Transport	Quality of local public transport
Q.Sport	Quality of local sport facilities
Q.Parks	Quality of local parks
Q.Commuinty	Quality of local community facilities
H.Qual	Housing Quality
Social	Socio-economic class

Other Abbreviations Used Within the Text

CHCS	Cardiff Housing Condition Survey
CPD	Central Postcode Directory
ED	Enumeration District
GIS	Geographic Information Systems
HCS Area	Housing Condition Survey Area
OSAPR	Ordnance Survey ADDRESS-POINT Reference
PAF	Postcode Address File
PED	Pseudo-Enumeration District
PIP	POINT-IN-POLYGON
UPRN	Unique Property Reference Number

Bibliography

Adair, A.S., Berry, J.N. and McGreal, W.S. (1996): Hedonic modelling, housing submarkets and residential valuation
Journal of Property Research. Vol 13. pp. 67-83

Alonso, W. (1964): **Location and Land Use.**
Harvard University Press, Cambridge, MA.

Anas, A. and Dendrinos, S.D. (1976): The new urban economics: A brief summary. in George J. Papageorgiou (ed): **Mathematical Land Use Theory.**
Lexington Books.

Anscombe, F.J. (1973): Graphs in statistical analysis.
The American Statistician. Vol. 27. pp. 17-22

Anselin, L. (1988a): **Spatial Econometrics: Methods and Models.**
Dordrecht: Kluwer Academic.

Anselin, L (1988b): Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity
Geographical Analysis. Vol. 20. No. 1. pp. 1-17

Anselin, L. (1992): Spatial dependence and spatial heterogeneity: model specification issues in the spatial expansion paradigm in E. Cassetti and J.P. Jones (eds): **Applications of the Expansion Method.**
London: Routledge 1992.

Anselin, L and Griffith, D.A. (1988): Do spatial effects really matter in regression analysis?
Papers of Regional Science Association. Vol 65. pp. 11-34

Bailey, T.C. (1994): A review of statistical spatial analysis in Geographical Information Systems in S. Fotheringham and P. Rogerson (eds): **Spatial Analysis and GIS**
Taylor and Francis Ltd. pp. 13-44.

Bajic, V. (1984): An analysis of the demand for housing attributes.
Applied Economics. Vol 16. 1984. pp. 597-610

Ball, M. (1973): Recent empirical work on the determinants of relative house prices.
Urban Studies. Vol. 10. 1973. pp. 213-233

Ball, M (1974): The determinants of relative house prices: A reply.
Urban Studies. Vol 11. 1974. pp. 231-233

Ball, M. (1985): The urban rent question.
Environment and Planning A. 1985. Vol. 17. No. 4. pp. 503-525.

Ball, M. and Kirwan, R. (1977): Accessibility and constraints in the urban housing market.

Urban Studies. Vol. 14. 1977. pp. 11-32.

Bassett, K. and Short, J. (1980): **Housing and Residential Structure: Alternative Approaches.**
Routledge and Kegan Paul: London.

Batty, M. (1976): **Urban Modelling: Algorithms, Calibrations, Predictions**
Cambridge University Press.

Batty, M. (1994). Urban models twenty five years on
Environment and Planning B. Vol. 21. No. 5. pp. 515-516

Batty, M. and Longley, P. (1987): A fractal based description of urban form
Environment and Planning B. Vol. 14. No. 2. pp. 123-134.

Batty, M. and Xie. Y.(1994a): Modelling inside a GIS 1: Model structures, exploratory data analysis and aggregation
International Journal of GIS. 1994. Vol. 8. No. 3. pp. 291-307

Batty, M. and Xie. Y.(1994b): Modelling inside a GIS 2: Selecting and calibrating urban models using Arc-Info
International Journal of GIS. 1994. Vol. 8. No. 5. pp. 451-470

Beaverstock, J. Leyshon, A. Rutherford, T. Thrift, N. and Williams, P. (1992): Moving houses: the geographical reorganisation of the estate agency industry in England and Wales in the 1980s.
Transactions, Institute of British Geographers. Vol 17. 1992 pp. 166-182

Beckman, M.J. (1968): **Location Theory**
Random House, Inc.

Belsley, D.A. and Kuh, E. and Welsch, R.E. (1980): Robust estimation of an hedonic housing price equation.
Regression Diagnostics. Chapter 4.4. pp 229-245.

Bender, B. and Hwang, H. (1985): Hedonic house price indices and secondary employment centres
Journal of Urban Economics. Vol. 17. pp. 90-107

Berry, B.J. (1976): Ghetto expansion and single family housing prices: Chicago 1968-1972
Journal of Urban Economics. Vol. 3. pp. 397-423

Blake, M and Openshaw, S. (1996):
School of Geography, Leeds University, Leeds. LS2 9JT
http://www.geog.leeds.ac.uk/staff/m.blake/v-sel/v-sel.htm

Boddy, M. (1980): **The Building Societies**
Macmillan, London.

Bourne, L.S. (1981): **The Geography of Housing.**
London: Edward Arnold

Boyle, P. and Dunn, C.E. (1991): Redefinition of Enumeration District centroids: a test of their accuracy by using Thiessen Polygons
Environment and Planning A. Vol. 23. pp. 1111-1119

Brown, G.M. and Pollakowski, H.O. (1977): The economic valuation of shore line
Review of Economics and Statistics. Vol. 59. pp. 272-278.

Brunsdon, C. Fotheringham, A.S. and Charlton, M.E. (1996): Geographically Weighted Regression: A Method of exploring spatial non-stationarity
Geographical Analysis. Vol. 28. No. 4. pp. 281-298.

Butler, R. (1982): The specification of hedonic indexes for urban housing.
Land Economics. Vol. 58. 1982. pp. 98-108

Can, A. (1990): The measurement of neighbourhood dynamics in urban house prices.
Economic Geography. Vol. 66. 1990. pp. 254-272.

Can, A. (1992): Specification and estimation of hedonic house price models.
Regional Science and Urban Economics. Vol. 22. 1992 pp. 453-474.

Cassetti, E. (1972): Generating models by the expansion method: Applications to geographical research
Geographical Analysis. Vol. 4. pp. 81-91

Cassetti, E. (1992): The dual expansion method: An application for evaluating the effect of population growth on development in E.Cassetti and J.P.Jones (eds): **Applications of the Expansion Method**
London: Routledge 1992

Chambers, J.M., Kleiner, B. and Tukey, P.A. (1983): **Graphical Methods For Data Analysis**
Duxbury Press, Boston. pp. 258-259

Chatterjee, S. and Price, B. (1977): **Regression Analysis by Example**
John Wiley and Sons

Cheshire, P. and Sheppard, S. (1995): On the price of land and the value of amenities
Econometrica. Vol. 62. No. 246. pp. 247-267

Cliff, A.D. and Ord, J.K. (1981): **Spatial Processes: Models and Applications.**
Pion, London.

Collins, A and Evans A (1994): Aircraft noise and residential property values: An artificial neural network approach
Journal of Transport Economics and Policy Vol. 27-28. pp. 175-198.

Cooley, R., Hobbs, M., and Clewer, A. (1995): Location and residential property values in M.Fischer, T. Sikos and L. Bassa (eds): **Recent Developments in Spatial Information, Modelling and Processing**
Geomarket Co.

Cox, K.R. (1979): **Location and Public Problems.**

Basil Blackwell: Oxford.

Cropper, M.L., Deck, L.B., McConnell, K.E. (1988): On choice of functional form for hedonic price functions
Review of Economics and Statistics. Vol. 70. NO. 4. pp. 668-675.

Cubin, J.S. (1970): A hedonic approach to some aspects of the Coventry housing market
Warwick Economic Research papers 14

Cushnie, J. (1994): A British standard is published
Mapping Awareness. Vol. 8. No. 5. pp. 40-43

Daniels, C.B. (1975): The influence of racial segregation on housing prices
Journal of Urban Economics. Vol 2. 1975. pp. 105-122.

Darling, A.H. (1973): Measuring benefits generated by urban water parks.
Land Economies Vol. 49. pp. 22-34.

Daultrey, S. (1976): **Principal Components Analysis**
Concepts and Techniques in Modern Geography 8
Norwich: Geo Books.

Daunton, M.J. (1977): **Coal Metropolis: Cardiff 1870-1914**

Davies, G. (1974): An econometric analysis of residential amenity
Urban Studies. Vol. 11. pp. 217-226

Dear, M.J. (1976): Spatial externalities and locational conflict in D. Massey and P. Batey (eds):
London Papers in Regional Science: Alternative Framework for Analysis
London, Pion. pp. 152-167

DeLisle, J.R. (1984): Market segmentation: implications for residential appraisal
The Real Estate Appraiser and Analyst. Vol. 50. pp. 48-54

Department of the Environment (1991): **The English Housing Condition Survey**
London: HMSO

Dinan, T.M. and Miranowski, J.A. (1989): Estimating the implicit price of energy efficiency. improvements in the residential housing market: A hedonic approach.
Journal of Urban Economics. Vol 25. 1989. pp. 52-67.

Dixon, T. (1992): I.T. Applications in property: An overview and future trends
Mapping Awareness Vol 6. No2.

Do, A.Q. and Grunditski,G. (1995): Golf-course and residential house prices - an empirical examination
Journal of Real Estate Finance and Economics. Vol. 10. No. 3, pp. 261-270

Do, A.Q., Wilbur, R. and Short, J. (1994): An empirical examination of the externalities of neighbourhood churches on housing values
Journal of Real Estate Finance and Economics. Vol. 9. pp. 127-136

Dorling, D. (1995): **A New Social Atlas of Britain**
John Wiley and Sons

Duban, R.A. and Sung, C. (1987): Spatial variation in the price of housing. Rent gradients in non-monocentric cities.
Urban Studies. Vol 24. 1987. pp 193-204.

Duban, R.A. and Sung, C. (1990): Specification of hedonic regressions: non-nested tests on measures of neighbourhood quality.
Journal of Urban Economics. Vol 27. 1990. pp. 97-110.

Dunn, R. (1989): Building regression models: the importance of graphics.
Journal of Geography in Higher Education. Vol. 13. No 1. 1989. pp. 15-30.

Dunn, R. Longley, P.A., and Wrigley, N. (1987): Graphical procedures for identifying functional form in binary discrete choice models - a case study of revealed tenure choice
Regional Science and Urban Economics, 1987. Vol. 17. No. 1. pp. 151-171

Ellickson, B. (1981): An alternative test of the hedonic theory of housing markets.
Journal of Urban Economics. Vol 9. 1981. pp. 56-79

Evans, A.W. (1973): **The Economics of Residential Location.**
London: Macmillan.

Fleming, M.C. and Nellis, J.G. (1981): The interpretation of house price statistics for the UK.
Environment and Planning A. Vol 13. pp 1109-1124.

Fleming, M.C. and Nellis, J.G. (1985): Research policy and review 2. House price statistics for the United Kingdom: A survey and critical review of recent developments
Environment and Planning A. Vol 17. pp 297-318

Fleming, M.C. and Nellis, J.G. (1987): **Spon's House Price Data Book.**
London: Spon.

Fleming, M.C. and Nellis, J.G. (1992): Development of standardised indices for measuring House price inflation incorporating physical and locational characteristics
Applied Economics. Vol 24. pp. 1067-1085.

Fleming, M.C. and Nellis, J.G. (1994):The measurement of UK house prices: A review and appraisal of the principal sources.
Housing Finance. No. 24

Flowerdew, R. (1991): Spatial data integration in D.J. Maguire, M.F. Goodchild and D.W. Rhind (eds): **Geographical Information Systems: Methodology and Potential Applications**
Longmans. London. pp.

Flowerdew, R. and Openshaw, S. (1987): A review of the problems of transferring data from one set of areal units to another incomplete set.

Research Paper 4. Northern Regional Research Laboratory, Lancaster and Newcastle.

Follain, J.R. and Jimenez, E. (1985): Estimating the demand for housing characteristics: A critique and survey.
Regional Science and Urban Economics. Vol 15. 1985. pp. 77-107.

Forrest, D., Glen, J. and Ward, R (1996) The impact of a light rail system on the structure of house prices - A hedonic longitudinal study
Journal of Transport Economics and Policy. Vol. 30. No. 1. pp. 15-37

Fortney, J (1996: A cost-benefit locational-allocation model for public facilities - An econometric approach
Geographical Analysis. Vol. 28. No. 1. pp. 67-92

Foster, C.D. and Jackman, R.A. and Perlman, M. (1980): **Local Government Finance in a Unitary State.**
George Allen and Unwin: London.

Foster, S.A. (1991) The expansion method: Implications for geographical research
Professional geographer. Vol. 43. No. 2. pp. 131-142

Fotheringham, A.S. and Wong, D.W.S (1991): The Modifiable Areal Unit Problem in multivariate statistical analysis.
Environment and Planning A. Vol. 23. No. 7 pp. 1025-1044

Fox, J. (1991): **Regression Diagnostics.**
Sage Publications

Freeman, A.M. (1979a): **The Benefit of Environmental Improvement: Theory and Practice**
John Hopkins University Press. Baltimore and London.

Freeman, A.M. (1979b): Hedonic prices, property values and measuring environmental benefits: A survey of the issues.
Scandinavian Journal of Economics. Vol. 81. pp. 154-173.

Freeman, R. and Dixon, T. (1992): **Information Technology Applications in Commercial Property**
MacMillian Press Ltd.

Garrod, G. and Willis, K. (1992a): Valuing the goods characteristics - an application of the hedonic price method to environmental attributes
Journal of Environmental Management. Vol. 34. No. 1 pp. 59-76

Garrod, G. and Willis, K. (1992b): The environmental economic impact of woodland: A two stage hedonic price model of the amenity value of forestry in Britain
Applied Economics. Vol. 24. pp. 715-728

Gatrell, A.C. (1989): On the spatial representation and accuracy of address-based data in the United Kingdom
International Journal of Geographical Information Systems. Vol 3. pp. 335-348

Gatrell, A.C., Dunn, C.E., and Boyle, P.J. (1991): The relative utility of the Central Postcode Directory and PinPoint Address Code in applications of Geographic Information Systems
Environment and Planning A. Vol. 23. pp. 1447-1458

Getis, A. (1994): Spatial dependence and heterogeneity and proximal databases in Fotheringham, S. and Rogerson, P. (eds): **Spatial Analysis and GIS**
Taylor and Francis Ltd. pp. 105-120

Gillard, Q. (1981): The effect of environmental amenities on house values: the example of a view lot
Professional Geographer Vol. 33. No. 2. 1981. pp. 216-220.

Goldstein, H. (1987): **Multi-level Models in Educational and Social Research**
Charles Griffin, London.

Goldstein, H. (1995): **Multi-level Statistical Models**
Kendals Library of Statistics 3.
Second Edition

Goodchild, M.F. (1986): **Spatial Autocorrelation**
Concepts and Techniques in Modern Geography 47
Norwich: Geo Books.

Goodchild, M.F. (1987): A spatial analytical perspective on Geographical Information Systems
International Journal of Geographical Information Systems. Vol. 1. pp. 335-354

Goodchild, M.F (1988): Towards an enumeration and classification of GIS in R.T. Aangeenburg and Y.M. Schiffman (eds): **International Geographic Information Systems (IGIS) Symposium: The Research Agenda. pp. 67-77**

Goodchild, M.F. (1991): The technological setting of GIS in D.J. Maguire, M.F. Goodchild and D.W.Rhind (eds): **Geographical Information Systems: Methodology and Potential Applications**
Longmans. London. pp. 45-54

Goodman, A.C. (1978): Hedonic prices, price indices and housing markets
Journal of Urban Economics. Vol. 5. pp. 471-484

Goodman, A.C. (1979): Externalities and non-monocentric price distance functions.
Land Economics. 1979. Vol 16. pp. 321-328.

Goodman, A.C. (1981): Housing submarkets within urban areas - definitions and evidence
Journal of Regional Science. Vol. 21. No. 2. pp. 175-185

Goodman, M. (1990): Players are hungry for a slice of cake
Money Weekly 7. Feb. pp. 9.

Gordon, P., Richardson, H.W. and Wong, H.L.(1986): The distribution of population and employment in a polycentric city: the case of Los Angeles.
Environment and Planning A. Vol. 18. pp. 161-173

Graves, P., Murdoch, J.C., Thayer, M.A. and Waldman, D. (1988): The robustness of the hedonic price function - urban air quality
Land Economics. Vol 64. No. 3. pp. 220-233

Grether, D.M. and Miesszkowski, P. (1980): The effects of non-residential landuses on the price of adjacent housing: some estimates of proximity effects
Journal of Urban Economics. Vol. 8. pp. 1-15

Griffith, D.A. (1981): Evaluating the transformation from a monocentric to a polycentric city
Professional Geographer. Vol. 33. No. 2. pp. 189-196

Griliches, Z. (1971): **Price Indexes and Quality Change.**
Harvard University Press.

Gross, D.J. (1988): Estimating willingness to pay for housing characteristics: An application of the Ellickson bid-rent model
Journal of Urban Economics. Vol 24. 1988. pp. 95-112.

Halvorsen, R. and Pollakowski, H. (1982): Choice of functional form for hedonic price equations
Journal of Urban Economics. Vol. 10. pp. 37-49.

Hamilton, B. (1982): Wasteful commuting
Journal of Political Economy. Vol. 90. pp. 1035-1053

Harrison, D.J. and Rubinfeld, D.L. (1978): Hedonic housing prices and the demand for clean air.
Journal of Environmental Economics and Management. Vol 5. 1978. pp. 81-102.

Harvey, D. (1972): Society, the city and the space-economy of urbanism.
Resource Paper No. 18. Association of American Geography. Washington D.C.

Harvey, D. (1973): **Social Justice and the City**
Edward Arnold Ltd.

Heikkilla, E. (1988): Multicollinearity in regression models with multiple distance measures
Journal of Regional Science. Vol. 28. pp. 345-361

Heikkilla, E., Gordon, P., Kim, J., Deiser, R.B. and Richardson, H.W. (1989): What happened to the CBD-distance gradient? Land values in a poliocentric city
Environment and Planning A. Vol. 21. pp. 221-232

Herrin, W. and Kern, C. (1992): Testing the standard urban model of residential choice: An implicit markets approach
Journal of Urban Economics. Vol. 31. pp. 145-163

Higgs, G. and Martin, D. (1995): **A Comparison of Recent Spatial Referencing Approaches in Planning**
Papers in Planning Research. No. 155
Department of City and Regional Planning. University of Cardiff

Hirschfield, A. , Brown, P. and Todd, P. (1995): GIS and the analysis of spatially referenced crime data: experiences in Merseyside, UK.
International Journal of Geographic Information Systems. Vol 9. No 2. pp 191-210

Hughes, W. and Sirmans, C.F. (1992): Traffic externalities and single-family house prices
Journal of Regional Science. Vol. 32. No. 4. pp. 487-500

Jackson, J.R. (1979): Intraurban variations in the price of housing.
Journal of Urban Economics. Vol 6. 1979. pp. 464-479.

Johnson, R.C and Kaserman, D.L. (1983): Housing market capitalization of energy-saving durable good investments
Economic Inquiry. Vol. 21. No. 3. pp. 374-386

Johnston, R.J., Gregory, D. and Smith, D.M. (1994): **The Dictionary of Human Geography**
Blackwell Publishers Ltd

Jones, K. (1991): **Multi-level Models for Geographical Research**
Concepts and Techniques in Modern Geography 54
Norwich: Geo Books

Jones, K. and Bullen, N. (1993): A Multilevel analysis of the variations in domestic property prices: Southern England 1980-1987.
Urban Studies. Vol 30. No. 8. 1993. pp. 1409-1426.

Jones, K. and Bullen, N. (1994): Contextual models of urban house prices: A comparison of fixed - and random- coefficient models developed by expansion.
Economic Geography. 1994. Vol. 70. No. 3. pp. 252-272

Kain, J.F. and Quigley, J.M. (1970a): Measuring the value of housing quality.
Journal of the American Statistical Association.

Kain, J.F. and Quigley, J.M. (1970b): Evaluating the quality of the residential environment.
Environment and Planning A Vol. 2. pp. 23-32

Keltics (1989): **The Cardiff Housing Condition Survey. Phase 1: Inner Area Final Report**
Keltics, Consulting Architects and Engineers Ltd.

Kennedy, P. (1985): **A Guide to Econometrics.**
Basil Blackwell. Second Edition.

Kennedy, G.A., Dai, M., Henning, S.A., and Vendeveer, L.R. (1996): A GIS-based approach for including topographic and locational attributes in the hedonic analysis of rural land values
American Journal of Agricultural Economics. Vol. 78. No. 5. pp. 1419-1437

Knox, P.L. (1995): **Urban Social Geography: An Introduction**
Routledge. Third edition

Krumm, R.J. (1980): Neighbourhood amenities: an economic analysis
Journal of Urban Economics. Vol. 7. pp. 208-224

Lake, I. (1996): Using Ordnance Survey digital map data to implement the hedonic pricing method: Assessing visual and road transport disamenity
University of East Anglia. School of Environmental Studies.

Lancaster, K. (1966): A new approach to consumer theory.
Journal of Political Economy. Vol. 74. 1966. pp 132-157.

Lansford, N.H. and Jones, L.L. (1995): Recreational and aesthetic value of water using hedonic price analysis
Journal of Agricultural and Resource Economics. Vol. 20. No. 2. pp. 341-355

Lerman, S.R. and Kern, C.R. (1983): Hedonic theory, bid-rent, and willingness-to-pay: some extensions of Ellicksons results.
Journal of Urban Economics. Vol 13. 1983. pp. 358-363.

Levesque, T. (1994): Modelling the effects of airport noise on residential housing markets
Journal of Transport Economics and Policy. Vol 28. No. 2. pp. 199-211

Li, M.M. and Brown, H.J. (1980): Micro-neighbourhood externalities and hedonic house prices
Land Economics. Vol. 56. pp. 125-140

Linnenman, P. (1980): Some empirical results on the nature of the hedonic price function for urban housing markets.
Journal of Urban Economics. Vol 8. pp. 47-68.

Longley, P.A. and Clarke, G. (1995): **GIS for Business and Service Planning**
Pearson Professional Ltd. 1995

Longley, P.A. and Dunn, R. (1988) Graphical assessment of housing market models
Urban Studies. 1988. Vol. 25. No. 1 pp. 21-33

Longley, P. and Higgs, G. and Martin, D. (1994): The predictive use of GIS to model property valuations.
International Journal of Geographical Information Systems. Vol. 8. No 2. pp. 217-235.

Longley, P. and Martin, D. and Higgs, G. (1993): The geographical implications of changing local taxation regimes.
Transactions. Institute of British Geographers. Vol 18. 1993. pp. 86-101.

MacDonald, J.F. (1979): **Economic Analysis of an Urban Housing Market.**
Studies in Urban Economics
Academic Press, 1979.

Maclennan, D. (1982): **Housing Economics: An Applied Approach.**
Longman: London and New York.

Maddala, G.S. (1992): **Introduction to Econometrics**
Academic Press, 1992. Second Edition.

Martin, D. (1992): Postcodes and the 1991 Census of population: issues, problems and prospects
Transactions of the Institute of British Geographers. Vol 17. pp. 350-357

Martin, D. (1996): **Geographic Information Systems and Their Socio-Economic Applications**
Routledge, London. Second Edition

Martin, D. and Higgs, G.(1994): **Georeferencing People and Places: a Comparison of Detailed Data Sets**
Papers in Planning Research.
Department of City and Regional Planning. University of Cardiff

Martin, D. and Longley, P. and Higgs, G. (1992): The geographical incidence of local government revenues: an intraurban case study.
Environment and Planning C. Vol 10. 1992. pp. 253-265.

Mather, P.M. (1976): **Computational Methods of Multivariate Analysis in Physical Geography**
John Wiley and Sons 1976

McLeod, P.B. (1984): The demand for local amenity: An hedonic price analysis
Environment and Planning A. Vol. 16. pp. 359-400.

McMillan, M. and Reid, B.G. and Gillen, D.W. (1980): An extension of the hedonic approach for estimating the value of quiet.
Land Economics. Vol 56. No 3. 1980. pp. 315-327.

Merrett, S. and Gary, F. (1982): **Owner Occupation in Britain.**
London: Routledge and Kegan Paul.

Millington, A.F. (1990): An Introduction to Property Valuation
London: Estates Gazette

Mills, E.S. (1972): **Urban Economics.**
Scott, Foresman and Company: London.

Mingche, M.L. and Brown, J.H. (1980): Micro-neighbourhood externalities and hedonic housing price.
Land Economics. Vol 56. 1980. pp. 125-141.

Munro, M. and Maclennan, D. (1987): Intra-urban changes in housing prices: Glasgow 1972-83.
Housing Studies. Vol 2. No 2. 1987. pp. 65-81.

Muth, R.F. (1969): **Cities and Housing: The Spatial Pattern of Urban Residential Landuse.**
The University of Chicago Press.

Muth, R.F. and Goodman, A.C. (1989) **The Economics of Housing Markets**
Fundamental and Applied Economics 31
Harwood Academic Publishers 1989

Nelson, J.P (1980): Airports and property values
Journal of Transport Economics and Policy. Vol 14. No. 1. pp. 37-52

Nicol, C. (1996): Interpretation and compatibility of house price series
Environment and Planning A. Vol 28. No. 1. pp. 119-133

North, J.H. and Griffin, C.C. (1993): Water sources as a housing characteristics - hedonic property valuation and willingness-to-pay for water
Water Resources Research. Vol. 29. No. 7. pp. 1923-1929

Odland, J. Golledge, R.G. and Rogerson, P.A (1989): Mathematical and statistical analysis in human geography in G.L. Gailey and J. Willmott (eds): **Geography in America**
Merrill Publishing Company. pp. 719-745

Ohsfeldt, R.L. (1988): Implicit markets and the demand for housing characteristics.
Regional Science and Urban Economics. Vol. 18. 1988. pp. 321-343.

Openshaw, S. (1984) **The Modifiable Areal Unit Problem**
Concepts and Techniques in Modern Geography 38
Norwich: Geo Books.

Openshaw, S (1991) Developing appropriate spatial analysis methods for GIS in D.J. Maguire, M.F. Goodchild and D.W.Rhind (eds): **Geographical Information Systems: Methodology and Potential Applications**
Longmans. London. pp. 389-402

Openshaw, S. (1994a) What is GISable spatial analysis?
New Tools for Spatial Analysis
Eurostat, Luxembourg. pp. 36-44

Openshaw, S. (1994b) Two exploratory space-time-attribute pattern analysers relevant to GIS. in Fotheringham, S. and Rogerson, P. (eds): **Spatial Analysis and GIS**
Taylor and Francis Ltd. pp. 83-104

Openshaw, S. (1995) Developing GIS relevant zone based spatial analysis methods
Paper at the Institute British Geographers Annual Conference. University of Newcastle. 1995

Openshaw, S. and Taylor. P.J. (1979): A million or so correlation coefficients: three experiments on the modifiable areal unit problem. in R.J. Bennett, N.J. Thrift, and N. Wrigley (eds): **Statistical Applications in the Spatial Sciences**
London: Pion.

Ordnance Survey (1993) **ADDRESS-POINT User Guide**
Ordnance Survey, Southampton

Ozanne, L. and Malpezzi, S. (1985): The efficacy of hedonic model estimation with the annual housing survey - evidence from demand experiments
Journal of Economic and Social Measurement. Vol. 13. No. 2. pp. 153-172

Palmquist, R.B. (1984): Estimating the demand for the characteristics of housing.
Review of Economics and Statistics. No. 66. pp. 394-404.

Papageorgiou, G.J. (1976): **Mathematical Landuse Theory**
Lexington Books.

Pinch, S. (1985): **Cities and services: The Geography of Collective Consumption.**
Routledge and Kegan Paul.

Pindyck, R.S. and Rubinfeld, D. (1991): **Econometric Models and Economic Forecasts.**
McGraw-Hill International Editions. Economic Series. 3rd Edition.

Powe, N.A., Garrod, D. and Willis, K.G. (1995): Valuation of urban amenities using an hedonic price model
Journal of Property Research. Vol. 12. pp. 137-147

Quigley, J.M. (1982): Non-linear budget constraints and consumer demand - an application to public programs for residential housing
Journal of Urban Economics. Vol. 12. No. 2. pp. 177-201

Raper, J., Rhind, D. and Shepard, J. (1992): **Postcodes: the New Geography**
Longman, London.

Rasbash, J. and Woodhouse, G. (1995): **Mln Command Reference**
London, Multilevel Models Project, Institute of Education.

Richardson, H.W. (1976): Relevance of mathematical land use theory to application. in George J. Papageorgiou (ed): **Mathematical Land Use Theory.**
Lexington Books.

Ridker, R. and Henning, J. (1968): The determination of residential property values with special reference to air pollution.
Review of Economics and Statistics. 1968.

Rogerson, P.A. and Fotheringham, A.S. (1994): GIS and spatial analysis: introduction and overview in Fotheringham, S. and Rogerson, P. (eds): **Spatial Analysis and GIS**
Taylor and Francis Ltd. pp. 1-10

Rosen, S. (1974): Hedonic prices and implicit markets: Product differentiation in pure competitions.
Journal of Political Economy. Vol. 72. 1974. pp. 34-55.

Sanchez. T.W. (1993): Measuring the impact of highways and roads on residential property values with GIS: An empirical analysis of Atlanta in Klosterman, R.E and French, S.P (eds): **Third International Conference on Computers in Urban Planning and Management. Vol. 2.**

Saunders, P. (1990): **A Nation of Home Owners.**
Unwin Hyman Ltd

Schnare, A. and Struyk, R.J. (1976): Segmentation in urban housing markets.
Journal of Urban Economics. Vol. 3. 1976. pp. 146-166.

Schroeder, L.D. and Sjoquist, D.L. and Stephan, P.E. (1986): **Understanding Regression Analysis: An Introductory Guide.**
Sage Publications 1986.

Short, J.R. (1982): **Housing in Britain: The Post-War Experience:**
Methuen.

Sinton, D.F (1992): Reflections on twenty five years of GIS
Insert in GIS World. February 1992. pp. 1-8

Smith, B. (1996): The NLIS in 1996: the pilot project expands
Mapping Awareness. Vol. 10. No. 2. pp. 22-24

Straszheim, M. (1974): Hedonic estimation of housing market prices. A further comment
Review of Economics and Statistics. pp. 404-406

Stull, W.J. (1975): Community environment, zoning, and the market value of single family homes
Journal of Law Economics. Vol. 18. pp. 535-558

Tyrvainen, L. (1997): The amenity value of the urban forest: An application of the hedonic pricing method
Landscape and Urban Planning. Vol. 37. No. 3. pp. 211-222

Veldhuisen, K.J. and Timmermans, H.P. (1984): Specification of individual residential utility-functions - a comparative analysis of three measurement procedures
Environment and Planning A. Vol. 16. No. 12. pp. 1573-1582

Wabe, J.S. (1971): A study of house prices as a means of estimating the value of journey time, the rate of time preferences and the valuation of some aspects of environment in the London Metropolitan Region.
Applied Economics. Vol. 3. pp. 247-255.

Waddell, P. and Berry, B.J.L (1993): Housing price gradients - The intersection of space and built form
Geographical Analysis. Vol. 25. No. 1. pp. 5-19

Waddell, P. and Berry, B.J.L. and Hoch, I. (1993): Residential property values in a multinodal urban area: New evidence on the implicit price of location
Journal of Real Estate Finance and Economics. Vol 7. 1993. pp. 117-141.

Wilkinson, R.K. (1971): The determinants of relative house prices.
Urban Studies. Vol. 8. 1971. pp. 50-59.

Wilkinson, R.K. (1973a): House prices and the measurement of externalities.
Economic Journal. Vol 83. 1973. pp. 72-86.

Wilkinson, R.K. (1973b): Measuring the determinants of relative house prices.

Environment and Planning. Vol 5. 1973. pp. 357-367.

Wilkinson, R.K. (1974): The determinants of relative house prices: a case of academic astigmatism.
Urban Studies. Vol. 11. 1974. pp. 227-230.

Witte, A. Sumka, J. and Erckson, H. (1979): An estimation of a structural hedonic price model of the housing market: An application of Rosen's theory of implicit markets
Econometrica. Vol. 47. pp. 1151-1173

Woodhouse, G., Rasbash, J.,Goldstein, H. and Yang, M. (1996): Introduction to multi-level modelling in Woodhouse, G. (eds) **Multi-Level Modelling Applications: A Guide for Users of Mln**
Institute of Education, University of London.

Wrigley, N. (1995): Revisiting the MAUP and the ecological fallacy in Cliff, A.D., Hoare, A.G. and Thrift, N.J. (eds): **Diffusing Geography**
Blackwell. 1995

Wyatt, P. (1994): Using a GIS for property valuation
Association of American Geographers. 1994. pp. 11-13.